

OPERATOR MEANS OF PROBABILITY MEASURES AND GENERALIZED KARCHER EQUATIONS

MIKLÓS PÁLFIA

ABSTRACT. In this article we consider means of positive bounded linear operators on a Hilbert space. We present a complete theory that provides a framework which extends the theory of the Karcher mean, its approximating matrix power means, and a large part of Kubo-Ando theory to arbitrary many variables, in fact, to the case of probability measures with bounded support on the cone of positive definite operators. This framework characterizes each operator mean extrinsically as unique solutions of generalized Karcher equations which are obtained by exchanging the matrix logarithm function in the Karcher equation to arbitrary operator monotone functions over the positive real half-line. If the underlying Hilbert space is finite dimensional, then these generalized Karcher equations are Riemannian gradients of convex combinations of strictly geodesically convex log-determinant divergence functions, hence these new means are the global minimizers of them, in analogue to the case of the Karcher mean as pointed out. Our framework is based on fundamental contraction results with respect to the Thompson metric, which provides us nonlinear contraction semigroups in the cone of positive definite operators that form a decreasing net approximating these operator means in the strong topology from above.

1. INTRODUCTION

Let E be Hilbert space and $S(E)$ denote the Banach space of bounded linear self-adjoint operators. Let $\mathbb{P} \subseteq S(E)$ denote the cone of positive definite operators on E . In this article we are concerned with means of members of \mathbb{P} that enjoy certain attractive properties that recently became important from the point of view of averaging in the finite dimensional case, see for example [4, 15, 18, 5, 6]. Usually the main difficulties here arise from the required property of operator monotonicity i.e., our means must be monotone with respect to the positive definite order on \mathbb{P} . The 2-variable theory of such functions is relatively well understood: each such function is represented by an operator monotone function according to the theory of Kubo-Ando [20]. However in the several variable case we have no such characterization of operator monotone functions.

Date: January 27, 2016.

2000 *Mathematics Subject Classification.* Primary 47A64, 46L05, Secondary 53C20, 53C35.

Key words and phrases. operator monotone function, operator mean, Karcher mean.

When E is finite dimensional, then there are additional geometrical structures on \mathbb{P} that are used to define n -variable operator means [3, 11, 26, 29]. In this setting \mathbb{P} is just the cone of positive definite n -by- n Hermitian matrices, where n is the dimension of E . It is a smooth manifold as an open subset of the vector space of n -by- n Hermitian matrices (which is just $S(E)$ in this case) and has a Riemannian symmetric space structure $\mathbb{P} \cong \mathrm{GL}(E)/\mathbb{U}$, where \mathbb{U} is the unitary group and $\mathrm{GL}(E)$ is the general linear group over E [9, 10]. This symmetric space is nonpositively curved, hence a unique minimizing geodesic between any two points exists [10]. The midpoint operation on this space, which is defined as taking the middle point of the geodesic connecting two points, is the geometric mean of two positive definite matrices [10]. The Riemannian distance on the manifold \mathbb{P} is of the form

$$d(A, B) = \sqrt{\mathrm{Tr} \log^2(A^{-1}B)}.$$

Then the (weighted) multivariable geometric mean or Karcher mean of the k -tuple $\mathbb{A} := (A_1, \dots, A_k) \in \mathbb{P}^k$ with respect to the positive probability vector $\omega := (w_1, \dots, w_k)$ is defined as the center of mass

$$(1) \quad \Lambda(w_1, \dots, w_k; A_1, \dots, A_k) = \arg \min_{X \in P(n, \mathbb{C})} \sum_{i=1}^k w_i d^2(X, A_i).$$

The Karcher mean was first considered in this setting in [29, 11]. The Karcher mean $\Lambda(\omega; \mathbb{A})$ is also the unique positive definite solution of the corresponding critical point equation called the Karcher equation

$$(2) \quad \sum_{i=1}^k w_i \log(X^{-1}A_i) = 0,$$

where the gradient of the function in the minimization problem (1) appears on the left hand side [29]. The (entry-wise) operator monotonicity of this mean in \mathbb{A} with respect to a fixed ω , that is $\mathbb{A} \leq \mathbb{B}$ if and only if $A_i \leq B_i$ for $1 \leq i \leq k$ implying $\Lambda(\omega; \mathbb{A}) \leq \Lambda(\omega; \mathbb{B})$ in the positive definite order, was unknown until it has been first proved in [26]. Later in [28] a one parameter family $P_s(\omega, \mathbb{A})$ of operator monotone means, called the matrix power means, has been constructed as the unique solution of the nonlinear operator equation

$$(3) \quad X = \sum_{i=1}^k w_i X \#_s A_i$$

for fixed $s \in [-1, 1]$, $w_i > 0$, $\sum_{i=1}^k w_i = 1$ and $A_i \in \mathbb{P}$, where

$$A \#_s B = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^s A^{1/2}$$

is the weighted two-variable geometric mean. Then in [28] it has been established that

$$\lim_{s \rightarrow 0+} P_s(\omega, \mathbb{A}) = \Lambda(\omega; \mathbb{A})$$

proving once more monotonicity and many other additional properties of Λ . Later the same approach based on the matrix power means has been adapted in [24], to extend $\Lambda(\omega; \mathbb{A})$ to the infinite dimensional setting as the unique solution of (2), even though there exists no Riemannian metric in this case, hence nor geodesically convex potential functions. These additional properties make the Karcher mean $\Lambda(\omega; \mathbb{A})$ the geometric mean to choose among the numerous different geometric means [11, 25]. From the geometric and analytic point of view the two characterizations provided by (1) and (2) may be the most significant.

In this paper we extend the theory of the Karcher mean and the approximating matrix power means appearing in [24, 25, 28] in three different ways. Firstly we generalize the setting of (2) by taking any operator monotone function $f : (0, \infty) \mapsto \mathbb{R}$ instead of just the special one $f(x) = \log(x)$. Here operator monotonicity of f is understood as $A \leq B$ implying $f(A) \leq f(B)$ for any $A, B \in \mathbb{P}$. The second generalization is that instead of sums for n-tuples of operators \mathbb{A} in (2) we consider integrals with respect to probability measures supported on \mathbb{P} . So our generalization of (2) has the form

$$(4) \quad \int_{\mathbb{P}} f(X^{-1}A) d\mu(A) = 0$$

for a probability measure μ on \mathbb{P} . Let \mathfrak{L} denote the set of all operator monotone functions $f : (0, \infty) \mapsto \mathbb{R}$ such that $f(1) = 0, f'(1) = 1$. Let $\mathcal{P}(\mathfrak{L} \times \mathbb{P})$ denote the set of Borel probability measures with bounded support on the product space $\mathfrak{L} \times \mathbb{P}$. Then our first result is the following.

Theorem 1.1. *Let $\mu \in \mathcal{P}(\mathfrak{L} \times \mathbb{P})$. Then the equation*

$$(5) \quad \int_{\mathfrak{L} \times \mathbb{P}} f\left(X^{-1/2}AX^{-1/2}\right) d\mu(f, A) = 0,$$

which we call the generalized Karcher equation, has a unique solution in \mathbb{P} that is also operator monotone in the extended sense of Definition 5.2 below, which provides a positive definite partial order for measures.

We prove this by suitably generalizing the argumentation in [24]. We establish a one parameter family of generalized matrix power means as fixed points of strict contractions, with suitable properties like operator monotonicity, such that they form a decreasing net in the partial positive definite order approximating the solution of (5) from above in the strong topology. Moreover these generalized matrix power means, called induced operator means, are unique solutions of (5) as well, hence fitting into the greater picture. By establishing many attractive properties, like operator monotonicity, of these fixed points, we prove that these properties are preserved in the limit i.e., a solution of (5) has these properties as well. Then we prove that the left hand side of (5) occurs as a Riemannian gradient of certain divergence functions on \mathbb{P} when E is finite dimensional, using the

integral formula

$$(6) \quad f(x) = \int_{[0,1]} l_s(x) d\nu(s)$$

valid for any $f \in \mathfrak{L}$, where $l_s(x) := \frac{x-1}{(1-s)x+s}$ and ν is a uniquely determined probability measure on $[0, 1]$. In particular we obtain a generalized form of (1) by means of strictly geodesically convex divergence functions

$$(7) \quad LD^s(X, A) = \frac{1}{s(1-s)} \operatorname{tr} \{ \log[(1-s)A + sX] - \log X \#_{1-s} A \}$$

where $s \in [0, 1]$, $X \#_{1-s} A$ is the weighted geometric mean of $X, A \in \mathbb{P}$. These divergence functions (7) also appeared recently in [14]. Let $\mathcal{P}([0, 1] \times \mathbb{P})$ denote the set of Borel probability measures on $[0, 1] \times \mathbb{P}$. Then our second main result is the following.

Theorem 1.2. *Let \mathbb{P} be finite dimensional and $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$. Then the minimization problem*

$$(8) \quad \arg \min_{X \in \mathbb{P}} \int_{[0,1] \times \mathbb{P}} LD^s(X, A) d\mu(s, A)$$

has a unique positive definite solution in \mathbb{P} . Moreover it is the unique solution of the corresponding generalized Karcher equation

$$\int_{[0,1] \times \mathbb{P}} X^{1/2} l_s(X^{-1/2} A X^{-1/2}) X^{1/2} d\mu(s, A) = 0,$$

which is the critical point equation of (8) with respect to the Riemannian metric of the symmetric space \mathbb{P} .

According to formula (6), the above theorem exhausts all possible generalized Karcher equations appearing in Theorem 1.1. Also the combination of Theorem 1.1 and Theorem 1.2, as a byproduct, yields a new property of operator monotone functions in \mathfrak{L} : they are characterized as Riemannian gradients of divergence functions.

Theorem 1.1 and Theorem 1.2 are extensions of the results known for the geometric (or Karcher) mean $\Lambda(\omega; \mathbb{A})$ [24, 25, 28, 29], in which case $f(x) = \log(x)$ in Theorem 1.1 and in Theorem 1.2 the measure $d\mu(s, A) = d(\nu \times \sigma)(s, A)$, where $d\nu(s) = ds$ and $d\sigma(A)$ is finitely supported over $A_i \in \mathbb{P}$ with corresponding weight w_i for $1 \leq i \leq k$. Also the matrix power means $P_s(\omega, \sigma)$ fit in the picture, in which case $f(x) = \frac{x^s-1}{s}$ in Theorem 1.1. Also our theorems extend the log-determinant means of positive matrices in [14]. Hence the unique solutions appearing in Theorem 1.1 and Theorem 1.2 not only provide us with new operator means, but unifies these existing approaches which - to our knowledge - is the first generally applicable extension of two-variable Kubo-Ando operator means [20] to several (noncommutative) variables and measures. Many nice properties of these generalized operator means are proved here including operator monotonicity, see Theorem 6.4 below.

The proof of Theorem 1.2 is based on a contraction principle in the cone \mathbb{P} with a complete metric space structure given by Thompson's part metric

$$d_\infty(A, B) = \max \{ \log M(A/B), M(B/A) \}$$

for any $A, B \in \mathbb{P}$, where $M(A/B) = \inf \{ \alpha : A \leq \alpha B \}$. The argument is an elaborate generalization of the one given in [24] which is motivated by the one in [28] based on the approximation of the Karcher mean with the matrix power means which are themselves fixed points of one parameter families of strict contractions. Also en route this process, new auxiliary results are proved for the Thompson metric as well which extend various earlier ones in [23].

The paper is organized as follows. In section 2 we introduce the divergence functions (7), calculate their gradients and prove their geodesic convexity along with new integral formulas for members in \mathfrak{L} . These results will provide Theorem 2.6 which gives the first half of Theorem 1.1. Then in section 3 we prove new integral formulas for operator means in the sense of Kubo-Ando. In section 4 we prove fundamental contraction results for operator means with respect to the Thompson metric and in section 5 we use these to build up an analogous theory of generalized matrix power means called induced operator means in our setting. Then in section 6 we prove Theorem 1.2 as a combined effort of Theorem 6.3, Theorem 6.9 and Theorem 6.13 which also concludes the second half of Theorem 1.1 and Theorem 1.2. In section 7 we briefly discuss the results obtained.

2. CONVEX CONE OF LOG-DETERMINANT DIVERGENCES OVER \mathbb{P}

Suppose now that the Hilbert space E is finite dimensional so that \mathbb{P} is isomorphic to the cone of positive definite matrices. In this case one can introduce the so called logarithmic-barrier function

$$f(X) = -\log \det X = -\operatorname{tr} \log X$$

which is strictly convex on \mathbb{P} [9]. Using this function we introduce the following one parameter family of divergence functions on \mathbb{P} also studied in [14]:

$$(9) \quad LD^s(X, A) = \frac{1}{s(1-s)} \operatorname{tr} \{ \log[(1-s)A + sX] - \log X \#_{1-s} A \}$$

for $s \in [0, 1]$ and $X, A \in \mathbb{P}$ where the cases $s = 0, 1$ are defined by a continuity argument and

$$(10) \quad X \#_{1-s} A = X^{1/2} \left(X^{-1/2} A X^{-1/2} \right)^{1-s} X^{1/2} = X (X^{-1} A)^{1-s}$$

is the weighted geometric mean. Indeed, one can check that

$$\begin{aligned} LD^1(X, A) &= \operatorname{tr} \{ X^{-1} A - I - \log(X^{-1} A) \} \\ LD^0(X, A) &= \operatorname{tr} \{ A^{-1} X - I - \log(A^{-1} X) \} \end{aligned}$$

by computing the limits $\lim_{s \rightarrow 1-} LD^s(X, A)$ and $\lim_{s \rightarrow 0+} LD^s(X, A)$. Moreover simple calculation show that

$$\begin{aligned} LD^s(X, A) &= \frac{1}{s(1-s)} \log \frac{\det[(1-s)A + sX]}{\det(X \#_{1-s} A)} \\ &= \frac{1}{s(1-s)} \operatorname{tr} \{ \log[(1-s)X^{-1}A + s] - \log(X^{-1}A)^{1-s} \}. \end{aligned}$$

The formulas defining $LD^s(X, A)$ show that indeed it is a *divergence function*, that is, it satisfies

- (I) $LD^s(X, A) \geq 0$,
- (II) $LD^s(X, A) = 0$ if and only if $X = A$,

where (I) essentially follows from the operator inequality between the arithmetic and geometric means

$$sX + (1-s)A \geq X \#_{1-s} A$$

and the operator monotonicity of \log , see for example [10]. Property (I) and (II) tells us that $LD^s(X, A)$ is a divergence function on \mathbb{P} . Now clearly any convex combinations of the form

$$\int_{[0,1]} LD^s(X, A) d\nu(s)$$

also satisfies (I) and (II) for any probability measure ν over the closed interval $[0, 1]$. Here we can consider arbitrary positive measures over $[0, 1]$, but we will see that the normalization condition $\int_{[0,1]} d\nu(s) = 1$ does not restrict our further analysis, since we will consider the point where these functions attain their minimum on \mathbb{P} .

We are interested in the convexity properties of $\int_{[0,1]} LD^s(X, A) d\nu(s)$. More precisely the geodesic convexity with respect to the Riemannian metric

$$(11) \quad \langle X, Y \rangle_A = \operatorname{tr} \{ A^{-1} X A^{-1} Y \}$$

where $A \in \mathbb{P}$ and $X, Y \in S(E)$, where $S(E)$ in this case is just the vector space of Hermitian matrices. It is well known that \mathbb{P} with the Riemannian metric is a Hadamard manifold i.e., it has nonpositive sectional curvature [10]. Therefore its geodesics connecting any two points $A, B \in \mathbb{P}$ are unique and are given by the geometric mean $A \#_t B$. The Levi-Civita connection with respect to the Riemannian metric is of the form

$$(12) \quad \nabla_{X_A} Y_A = DY[A][X_A] - \frac{1}{2} (X_A A^{-1} Y_A + Y_A A^{-1} X_A)$$

see for example [22].

Let us consider the set of operator monotone functions $f : (0, \infty) \mapsto \mathbb{R}$. Operator monotonicity means that for any two Hermitian matrices $X \leq Y$ we have $f(X) \leq f(Y)$. Such functions have strong analytic properties, each such function has a unique integral representation [9] of the form

$$(13) \quad f(x) = \alpha + \beta x + \int_0^\infty \frac{\lambda}{\lambda^2 + 1} - \frac{1}{\lambda + x} d\mu(\lambda),$$

where α is a real number, $\beta \geq 0$ and μ is a unique positive measure on $[0, \infty)$ such that

$$(14) \quad \int_0^\infty \frac{1}{\lambda^2 + 1} d\mu(\lambda) < \infty.$$

Notice that the set of all such functions is a convex cone. A subset of this cone will be of our interest:

$$\mathfrak{L} := \{f : (0, \infty) \mapsto \mathbb{R}, f \text{ is operator monotone}, f(1) = 0, f'(1) = 1\}.$$

The following is a simple corollary of Bendat and Sherman's result [8], see Theorem 3.7 in [16].

Corollary 2.1. *The function $f \in \mathfrak{L}$ if and only if there exists a unique probability measure ν on $[0, 1]$ such that*

$$(15) \quad f(x) = \int_{[0,1]} \frac{x-1}{(1-s)x+s} d\nu(s).$$

We can also consider

$$\begin{aligned} \int_{[0,1]} \frac{x-1}{(1-s)x+s} d\nu(s) &= \int_{[0,\infty]} \frac{x-1}{\frac{x}{\lambda+1} + \frac{\lambda}{\lambda+1}} d\nu\left(\frac{\lambda}{\lambda+1}\right) \\ &= \int_{[0,\infty]} \lambda + 1 - \frac{(\lambda+1)^2}{\lambda+x} d\nu\left(\frac{\lambda}{\lambda+1}\right) \end{aligned}$$

with $d\eta(\lambda) := d\nu\left(\frac{\lambda}{\lambda+1}\right)$, so that we obtain the following similar corollary as well:

Corollary 2.2. *The function $f \in \mathfrak{L}$ if and only if*

$$(16) \quad f(x) = \int_{[0,\infty]} \lambda + 1 - \frac{(\lambda+1)^2}{\lambda+x} d\eta(\lambda)$$

where η is a unique probability measure on the closed, compact interval $[0, \infty]$.

The above integral characterizations will be useful tools also in later sections. One can easily see that

$$(17) \quad 1 - x^{-1} \leq \frac{x-1}{(1-s)x+s} \leq x-1$$

holds for all $s \in [0, 1]$ and $x > 0$, by comparing the functions on the disjoint intervals $(0, 1)$ and $[1, \infty)$. Hence after integration we conclude that

$$(18) \quad 1 - x^{-1} \leq f(x) \leq x - 1$$

holds for all $f \in \mathfrak{L}$. Now we give yet another characterization of the set \mathfrak{L} .

Definition 2.1. Let \mathfrak{D}_A denote the simplex of divergence functions on \mathbb{P} of the form

$$F(X) = \int_{[0,1]} LD^s(X, A) d\nu(s)$$

for any probability measure ν over the closed interval $[0, 1]$ and $A \in \mathbb{P}$.

Note that the above integral exists, since for fixed $X, A \in \mathbb{P}$ the real function $s \mapsto LD^s(X, A)$ is continuous over the compact interval $[0, 1]$, hence bounded and strongly measurable so it is integrable by dominated convergence theorem or by Theorem 11.8 in [2]. The same reasoning ensures the existence of other integrals over $[0, 1]$ that are considered in this section.

Theorem 2.3. *Let $F \in \mathfrak{D}_A$ be represented by a probability measure ν over the closed interval $[0, 1]$. Then the Riemannian gradient is*

$$(19) \quad \begin{aligned} \nabla F(X) &= - \int_{[0,1]} X(X^{-1}A - I) [(1-s)X^{-1}A + sI]^{-1} d\nu(s) \\ &= -Xf(X^{-1}A) = -X^{1/2}f(X^{-1/2}AX^{-1/2})X^{1/2} \end{aligned}$$

with respect to the Riemannian metric (11) and $f \in \mathfrak{L}$ represented by ν as in (15).

Proof. The Riemannian gradient $\nabla F(X)$ of $F(X)$ is defined by the relation

$$\left. \frac{\partial}{\partial t} F(X + Vt) \right|_{t=0} = \langle \nabla F(X), V \rangle_X$$

for any V in the tangent space at X . I.e., we compute

$$\begin{aligned} \frac{\partial}{\partial t} LD^s(X + Vt, A)|_{t=0} &= \\ &= \frac{1}{s(1-s)} \operatorname{tr} \{ D \log((1-s)X^{-1}A + sI) (-(1-s)X^{-1}VX^{-1}A) \\ &\quad - (1-s)D \log(X^{-1}A) (-X^{-1}VX^{-1}A) \} \end{aligned}$$

where $D \log : \mathbb{P} \times S(E) \rightarrow S(E)$ is the Fréchet derivative of \log which is linear in the second variable. Moreover since \log is an analytic function on $(0, \infty)$ we have by the linearity, cyclic property of the trace and the Riesz-Dunford functional calculus [41], that for any X, Y we have that

$$\operatorname{tr} \{ D \log(X)(Y) \} = \operatorname{tr} \{ D \log(X)(I)Y \}.$$

Further calculation with this gives that

$$\begin{aligned} \frac{\partial}{\partial t} LD^s(X + Vt, A)|_{t=0} &= \\ &= \frac{1}{s} \operatorname{tr} \{ D \log(X^{-1}A)(I) X^{-1}VX^{-1}A \\ &\quad - D \log((1-s)X^{-1}A + sI)(I) X^{-1}VX^{-1}A \} \\ &= \frac{1}{s} \operatorname{tr} \left\{ A^{-1}XX^{-1}VX^{-1}A - [(1-s)X^{-1}A + sI]^{-1} X^{-1}VX^{-1}A \right\} \\ &= \frac{1}{s} \operatorname{tr} \left\{ VX^{-1} - X^{-1}A [(1-s)X^{-1}A + sI]^{-1} X^{-1}V \right\}, \end{aligned}$$

hence

$$\nabla F(X) = \frac{1}{s} X \left\{ I - [(1-s)I + s(X^{-1}A)^{-1}]^{-1} \right\}.$$

Now simple calculation shows that

$$\frac{1 - [1 - s + sx^{-1}]^{-1}}{s} = -\frac{x - 1}{(1 - s)x + s},$$

and that the whole calculation also holds even if $s = 0$ or 1 . Since Bochner integration is exchangeable with the linear differential operator of Fréchet differentiation, we get from the above that (19) holds. \square

Remark 2.1. The proof of Theorem 2.3 also shows that actually every $f \in \mathfrak{L}$ corresponds to a Riemannian gradient of a divergence function in \mathfrak{D}_A .

We say that a map $f : \mathbb{P} \rightarrow \mathbb{R}$ is geodesically convex with respect to the metric (11) if it is convex along any geodesic $\gamma : [0, 1] \rightarrow \mathbb{P}$ i.e., the function $f(\gamma(t))$ is convex:

$$f(\gamma(t)) \leq (1 - t)f(\gamma(0)) + tf(\gamma(1))$$

for $t \in [0, 1]$. Similarly f is strictly geodesically convex if additionally

$$f(\gamma(t)) < (1 - t)f(\gamma(0)) + tf(\gamma(1))$$

for $t \in [0, 1]$. It is well known that f is geodesically convex if and only if the Riemannian Hessian $\text{Hess } f(X) : S(E) \times S(E) \rightarrow \mathbb{R}$ is positive semidefinite, similarly f is strictly geodesically convex if and only if $\text{Hess } f(X)(\cdot, \cdot)$ is positive definite [19, 35]. Moreover if one can show that $\text{Hess } f(X) > m > 0$ on some bounded geodesically convex set, then it follows that f is uniformly convex i.e.,

$$f(\gamma(t)) \leq (1 - t)f(\gamma(0)) + tf(\gamma(1)) - \frac{m}{2}t(1 - t)d(\gamma(0), \gamma(1))^2$$

for all geodesics γ lying entirely in the bounded geodesically convex set. In the above inequality d is the Riemannian distance function.

Theorem 2.4. *Every $F \in \mathfrak{D}_A$ is strictly geodesically convex function with respect to the metric (11).*

Proof. The idea is to show that the Riemannian Hessian of F is positive definite i.e., $\text{Hess } F(X)(V, V) > 0$ for any nonzero $V \in S(E)$ in the tangent space at X . The Riemannian Hessian of F is defined by

$$\text{Hess } F(X)(V, W) = \langle \nabla_V(\nabla F(X)), W \rangle_X$$

as a bilinear form acting on the tangent space at X , where ∇_V is the covariant derivative given by the Levi-Civita connection (12), see for example [37]. By Theorem 2.3 we have that $\nabla F(X) = -Xf(X^{-1}A)$ with an $f \in \mathfrak{L}$, so we have

$$\begin{aligned} \nabla_V(\nabla F(X)) &= -Vf(X^{-1}A) + XDf(X^{-1}A)(X^{-1}VX^{-1}A) \\ &\quad + \frac{1}{2} [Xf(X^{-1}A)X^{-1}V + Vf(X^{-1}A)]. \end{aligned}$$

Hence we have

$$\begin{aligned} \text{Hess } F(X)(V, V) &= \langle \nabla_V(\nabla F(X)), V \rangle_X \\ &= \text{tr} \left\{ X^{-1} \left\{ -Vf(X^{-1}A) + XDf(X^{-1}A)(X^{-1}VX^{-1}A) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} [Xf(X^{-1}A)X^{-1}V + Vf(X^{-1}A)] \right\} X^{-1}V \right\}. \end{aligned}$$

By the cyclic property of the trace the above is equivalent to

$$\text{Hess } F(X)(V, V) = \text{tr} \{ Df(X^{-1}A)(X^{-1}VX^{-1}A)X^{-1}V \}.$$

Now we make use of the integral representation for f given in Corollary 2.2 in the form

$$f(x) = \int_{[0, \infty]} \lambda + 1 - \frac{(\lambda + 1)^2}{\lambda + x} d\eta(\lambda).$$

By exchanging the Fréchet derivative with the integral we obtain that

$$\begin{aligned} Df(X^{-1}A)(X^{-1}VX^{-1}A) &= \nu(\{\infty\})X^{-1}VX^{-1}A \\ &\quad + \int_{[0, \infty)} (\lambda I + X^{-1}A)^{-1}X^{-1}VX^{-1}A(\lambda I + X^{-1}A)^{-1}d\eta(\lambda), \end{aligned}$$

where we separated the term corresponding to $\{\infty\}$. Hence by the linearity of the trace we have

$$\begin{aligned} \text{Hess } F(X)(V, V) &= \nu(\{\infty\}) \text{tr} \{ X^{-1}VX^{-1}AX^{-1}V \} \\ &\quad + \int_{[0, \infty)} \text{tr} \{ (\lambda I + X^{-1}A)^{-1}X^{-1}VX^{-1}A(\lambda I + X^{-1}A)^{-1}X^{-1}V \} d\eta(\lambda). \end{aligned}$$

Hence it suffices to prove now that the expressions

$$\begin{aligned} c(V) &:= \text{tr} \{ X^{-1}VX^{-1}AX^{-1}V \} > 0, \\ p(V) &:= \text{tr} \{ (\lambda I + X^{-1}A)^{-1}X^{-1}VX^{-1}A(\lambda I + X^{-1}A)^{-1}X^{-1}V \} > 0 \end{aligned}$$

for all nonzero $V \in S(E)$, $X, A \in \mathbb{P}$ and $\lambda \geq 0$. Using the notation $P = X^{-1/2}AX^{-1/2}$, $H = X^{-1/2}VX^{-1/2}$, the cyclic property of the trace and that $(\lambda I + P)^{-1/2}$ commutes with $P^{1/2}$, we get that

$$\begin{aligned} p(V) &= \text{tr} \{ (\lambda I + P)^{-1}HP(\lambda I + P)^{-1}H \} \\ &= \text{tr} \left\{ (\lambda I + P)^{-1/2}H(\lambda I + P)^{-1/2}P^{1/2} \right. \\ &\quad \left. P^{1/2}(\lambda I + P)^{-1/2}H(\lambda I + P)^{-1/2} \right\} \\ &= \left\| (\lambda I + P)^{-1/2}H(\lambda I + P)^{-1/2}P^{1/2} \right\|_2^2, \\ c(V) &= \text{tr} \{ HPH \} \\ &= \text{tr} \left\{ HP^{1/2}P^{1/2}H \right\} \\ &= \left\| HP^{1/2} \right\|_2^2 \end{aligned}$$

where $\|T\|_2 = \sqrt{\text{tr}\{TT^*\}}$ is the Frobenius or Hilbert-Schmidt norm of $T \in \text{End}(E)$. Since $(\lambda I + P)^{-1/2}$ and $P^{1/2}$ are nonsingular, moreover H is nonzero, it follows that

$$\begin{aligned} \left\| (\lambda I + P)^{-1/2} H (\lambda I + P)^{-1/2} P^{1/2} \right\|_2^2 &> 0, \\ \left\| H P^{1/2} \right\|_2^2 &> 0 \end{aligned}$$

hence $p(V) > 0$ and $c(V) > 0$ which yields $\text{Hess } F(X)(V, V) > 0$ since η is a positive measure. \square

Theorem 2.4 has many important consequences as we will see shortly. Before that let us investigate under what circumstances is the Hessian bounded away from zero in the previous Theorem 2.4. By simple arguments one can see that

$$\begin{aligned} \left\| (\lambda I + P)^{-1/2} H (\lambda I + P)^{-1/2} P^{1/2} \right\|_2^2 &> \min\{1/(\lambda + p_i)\} \min\{p_i^{1/2}\} \|H\|_2^2 \\ \left\| H P^{1/2} \right\|_2^2 &> \min\{p_i^{1/2}\} \|H\|_2^2 \end{aligned}$$

where p_i are the eigenvalues of P . If λ is finite, and eigenvalues of A, X are bounded i.e., X, A are in a bounded metric ball and the measure η is supported over some closed interval $[a, b] \subset [0, \infty]$, then the above bounds on the right are strictly greater than 0 for unit length directions V , moreover can be chosen uniformly over the bounded set and the interval $[a, b]$. This shows the following.

Corollary 2.5. *The Hessian $\text{Hess } F(X)(\cdot, \cdot)$ in Theorem 2.4 has eigenvalues greater than some $c > 0$ on geodesically bounded sets i.e., every $F \in \mathfrak{D}_A$ is a uniformly geodesically convex function with respect to the metric (11) on bounded metric balls.*

Now Theorem 2.4 and Corollary 2.5 implies the following result. Let $\mathcal{P}(\mathbb{P})$ denote the set of all Radon probability measures over \mathbb{P} with bounded support in \mathbb{P} . The integrals in the following theorem exist by the continuity and boundedness of the integrands over the support of the measures, since continuity ensures strong measurability, while boundedness (which follows from continuity over the compact support of the measures) ensures integrability using the dominated convergence theorem or Theorem 11.8 in [2].

Theorem 2.6. *Let $\sigma \in \mathcal{P}(\mathbb{P})$ and let ν be a probability measure on $[0, 1]$. Let $C \subseteq \mathbb{P}$ be a closed, bounded geodesically convex set. Then*

I. *The solution of the optimization problem*

$$(20) \quad \min_{X \in C} \int_{\mathbb{P}} \int_{[0,1]} LD^s(X, A) d\nu(s) d\sigma(A)$$

exists and is unique in C .

II. *If the global optimization problem*

$$(21) \quad \min_{X \in \mathbb{P}} \int_{\mathbb{P}} \int_{[0,1]} LD^s(X, A) d\nu(s) d\sigma(A)$$

has a solution, then it is unique and satisfies the nonlinear operator equation

$$(22) \quad \int_{\mathbb{P}} X^{1/2} f(X^{-1/2} A X^{-1/2}) X^{1/2} d\sigma(A) = 0$$

where $f(x) = \int_{[0,1]} \frac{x-1}{(1-s)x+s} d\nu(s)$ and $f \in \mathfrak{L}$.

Proof. The uniqueness of the solution of the global problem follows from the strict convexity of $LD^s(X, A)$ in Theorem 2.4. Moreover if the minimizer exists then it is a critical point of the gradient of the objective function which is explicitly calculated in Theorem 2.3.

The existence of the unique solution of the local problem follows from the continuity and uniform convexity of $LD^s(X, A)$ in bounded geodesically convex sets according to Corollary 2.5 and a standard optimization argument for strongly convex functions bounded from below, see for example Theorem 1.7 in [39]. \square

Theorem 2.6 introduces the important notions that will be investigated in the remaining sections of the paper. The gradient equation (22) is of fundamental importance in this paper.

Definition 2.2 (Generalized Karcher equation). Let $f \in \mathfrak{L}$ and $\sigma \in \mathcal{P}(\mathbb{P})$. Then the nonlinear operator equation

$$\int_{\mathbb{P}} X^{1/2} f(X^{-1/2} A X^{-1/2}) X^{1/2} d\sigma(A) = 0$$

for $X \in \mathbb{P}$ is called the generalized Karcher equation for the function f .

Theorem 2.6 is just not strong enough for our purposes to show the existence of the global minimizer

$$\arg \min_{X \in \mathbb{P}} \int_{\mathbb{P}} \int_{[0,1]} LD^s(X, A) d\nu(s) d\sigma(A),$$

nor to show additional important properties satisfied by the unique solutions. The rest of this paper will be devoted to this existence problem, along with establishing the properties enjoyed by the solution, which are finally concluded in Theorem 7.1. We will study this problem indirectly by instead looking at the generalized Karcher equation

$$\int_{\mathbb{P}} X^{1/2} f(X^{-1/2} A X^{-1/2}) X^{1/2} d\sigma(A) = 0$$

corresponding to the minimization problem and show the existence of the unique solution of it in the more general setting when \mathbb{P} is possibly over an infinite dimensional Hilbert space E . This approach requires us to build a

widely applicable machinery which relies on principles of nonlinear contraction semigroups in the ordered cone \mathbb{P} .

From now on we abandon the finite dimensional case and let E be infinite dimensional so that \mathbb{P} is the full cone of positive definite operators.

3. OPERATOR MEANS AND REPRESENTATIONS

Let us recall the definition of operator (or matrix) mean from [20]:

Definition 3.1. A two-variable function $M: \mathbb{P} \times \mathbb{P} \mapsto \mathbb{P}$ is called a matrix or operator mean if

- (i) $M(I, I) = I$ where I denotes the identity,
- (ii) if $A \leq A'$ and $B \leq B'$, then $M(A, B) \leq M(A', B')$,
- (iii) $CM(A, B)C \leq M(CAC, CBC)$ for all Hermitian C ,
- (iv) if $A_n \downarrow A$ and $B_n \downarrow B$ then $M(A_n, B_n) \downarrow M(A, B)$,

where \downarrow denotes the convergence in the strong operator topology of a monotone decreasing net.

In property (ii), (iii), (iv) the partial order being used is the positive definite order i.e., $A \leq B$ if and only if $B - A$ is positive semidefinite. An important consequence of these properties is [20] that every operator mean can be uniquely represented by a positive, normalized, operator monotone function $f(t)$ in the following form

$$(23) \quad M(A, B) = A^{1/2} f\left(A^{-1/2} B A^{-1/2}\right) A^{1/2}.$$

This unique $f(t)$ is said to be the representing function of the operator mean $M(A, B)$. So actually operator means are in one-to-one correspondence with normalized operator monotone functions, the above characterization provides an order-isomorphism between them. Normalization means that $f(1) = 1$. For symmetric means i.e., for means $M(A, B) = M(B, A)$, we have $f(t) = t f(1/t)$ which implies that $f'(1) = 1/2$. Operator monotone functions have strong continuity properties, namely all of them are analytic functions and can be analytically continued to the upper complex half-plane.

The set of all operator means is denoted by \mathfrak{M} i.e.,

$$\mathfrak{M} = \{M(\cdot, \cdot) : M(A, B) = A^{1/2} f(A^{-1/2} B A^{-1/2}) A^{1/2}, f : (0, \infty) \mapsto (0, \infty) \text{ operator monotone, } f(1) = 1\}.$$

Similarly $\mathfrak{m} = \{f(x) : f \text{ is a representing function of an } M \in \mathfrak{M}\}$. We will use the notation $\mathfrak{m}(t)$ to denote the set of all operator monotone functions f on $(0, \infty)$ such that $f(x) > 0$ for all $x \in (0, \infty)$ and $f(1) = 1, f'(1) = t$. We can find the minimal and maximal elements of $\mathfrak{m}(t)$ for all $t \in (0, 1)$ easily.

Lemma 3.1. For all $f(x) \in \mathfrak{m}(t)$ we have

$$(24) \quad ((1-t) + tx^{-1})^{-1} \leq f(x) \leq (1-t) + tx.$$

Proof. Since every operator monotone function is operator concave, see Chapter V [9], therefore we must have $f(x) \leq (1-t) + tx$ by concavity and the normalization conditions on elements of $\mathfrak{m}(t)$. Since the map x^{-1} is order reversing on positive matrices, we have that if $f(x) \in \mathfrak{m}(t)$ then also $f(x^{-1})^{-1} \in \mathfrak{m}(t)$, which leads to the lower bound. \square

Since $((1-t) + tx^{-1})^{-1}$ and $(1-t) + tx$ are operator monotone, they are the minimal and maximal elements of $\mathfrak{m}(t)$ respectively, and also they are the representing functions of the weighted harmonic and arithmetic means. In general by the previous Lemma 3.1 the inequality

$$(25) \quad [(1-t)A^{-1} + tB^{-1}]^{-1} \leq M(A, B) \leq (1-t)A + tB.$$

is true for all $M(A, B)$ operator means with representing operator monotone function f for which we have $f'(1) = t$. In this sense $\mathfrak{m}(t)$ characterizes weighted operator means. If we take this as the definition of weighted operator means, one can compare it with the definition of weighted matrix means given in [34].

We will make use of the following characterization due to Hansen using the notation

$$h_s(x) := [(1-s) + sx^{-1}]^{-1}$$

for $s \in [0, 1]$.

Proposition 3.2 (Theorem 4.9 in [16]). *Let $f \in \mathfrak{m}$. Then*

$$(26) \quad f(x) = \int_{[0,1]} [(1-s) + sx^{-1}]^{-1} d\nu(s)$$

where ν is a probability measure over the closed interval $[0, 1]$.

There are two degenerate cases of operator means induced by a ν which are supported only over the single points 0 or 1. One of them is the left trivial mean

$$l(x) = 1$$

with represented operator mean $M(A, B) = A$ and the right trivial mean

$$r(x) = x$$

with represented operator mean $M(A, B) = B$.

Proposition 3.3. *Let Φ be a positive unital linear map and $M \in \mathfrak{M}$. Then*

$$\Phi(M(A, B)) \leq M(\Phi(A), \Phi(B))$$

for $A, B > 0$.

Proof. Using Proposition 3.2 we have that

$$M(A, B) = \int_{[0,1]} [(1-s)A^{-1} + sB^{-1}]^{-1} d\nu(s)$$

where ν is a probability measure on $[0, 1]$. By Theorem 4.1.5 in [10] we have that

$$\Phi([(1-s)A^{-1} + sB^{-1}]^{-1}) \leq [(1-s)\Phi(A)^{-1} + s\Phi(B)^{-1}]^{-1}.$$

Using the fact that ν can be approximated by finitely supported measures and the linearity of Φ , we get from the above that

$$\begin{aligned} \Phi\left(\int_{[0,1]} [(1-s)A^{-1} + sB^{-1}]^{-1} d\nu(s)\right) &= \int_{[0,1]} \Phi([(1-s)A^{-1} + sB^{-1}]^{-1}) d\nu(s) \\ &\leq \int_{[0,1]} [(1-s)\Phi(A)^{-1} + s\Phi(B)^{-1}]^{-1} d\nu(s). \end{aligned}$$

□

In [20] Kubo and Ando defined the transpose of a matrix mean $M(A, B)$ as

$$(27) \quad M'(A, B) = M(B, A).$$

By Proposition 3.2 it is clear that for an

$$M(A, B) = \int_{[0,1]} [(1-s)A^{-1} + sB^{-1}]^{-1} d\nu(s)$$

we have that

$$M'(A, B) = M(B, A) = \int_{[0,1]} [(1-s)B^{-1} + sA^{-1}]^{-1} d\nu(s).$$

So if $M'(A, B)$ has corresponding measure ν' , then $d\nu'(s) = d\nu(1-s)$. Similarly for the representing functions we have $f^{tr}(x) = xf(1/x)$. Also symmetric means $M(A, B) = M'(A, B)$ have corresponding probability measures ν such that $d\nu(s) = d\nu(1-s)$ and vice versa.

We have one more result characterizing the partial order between operator means which should be well known, however its proof cannot be found anywhere in the literature.

Proposition 3.4. *Let $M, N \in \mathfrak{M}$ with representing functions $f, g \in \mathfrak{m}$ respectively. Then $M(A, B) \leq N(A, B)$ for all $A, B \in \mathbb{P}$ if and only if $f(x) \leq g(x)$ for $x > 0$.*

Proof. From the definition of the transpose mean we have the following formulas

$$\begin{aligned} M(A, B) &= A^{1/2} f(A^{-1/2} B A^{-1/2}) A^{1/2} = M'(B, A) = B^{1/2} f^{tr}(B^{-1/2} A B^{-1/2}) B^{1/2} \\ N(A, B) &= A^{1/2} g(A^{-1/2} B A^{-1/2}) A^{1/2} = N'(B, A) = B^{1/2} g^{tr}(B^{-1/2} A B^{-1/2}) B^{1/2} \end{aligned}$$

where $f^{tr}(x) = xf(1/x)$ and $g^{tr}(x) = xg(1/x)$.

So now suppose $f(x) \leq g(x)$ for $x > 0$. Then it follows that also $f^{tr}(x) \leq g^{tr}(x)$, hence by the above correspondence $M(A, B) \leq N(A, B)$. The reverse implication also follows from the reverse of the same argument. □

4. CONTRACTION PROPERTY OF OPERATOR MEANS

In the section we prove further properties of operator means using explicitly the integral characterizations given in Proposition 3.2. We will use the results given in this section to generalize the construction of matrix power means to all possible operator means in later sections.

Let E be a Hilbert space, $\mathfrak{B}(E)$ denote the Banach space of bounded linear operators, $S(E)$ denote the Banach space of bounded linear self-adjoint operators and $\mathbb{P} \subseteq S(E)$ the cone of positive definite and \mathbb{P}_0 the cone of positive semi-definite operators. On \mathbb{P} we have the positive definite order similarly to the finite dimensional case which means that $A \leq B$ for $A, B \in \mathbb{P}$ if and only if $B - A \in \mathbb{P}_0$. It is also easy to see that if for $A, B \in S(E)$ and $0 \leq A \leq B$ then also $\|A\| \leq \|B\|$ [24]. We will use the notation $[A, B]$ for the order interval generated by $A \leq B$ i.e., $[A, B] = \{X \in \mathbb{P} : A \leq X \leq B\}$. We also have that $\mathbb{P} = \bigcup_{k=1}^{\infty} [\frac{1}{k}I, kI]$.

We would also like to consider measures taking values in \mathbb{P} and we would like to integrate with respect to these measures. There are a number of different (stronger or weaker) ways to do that. In our setting the integral that we need is the weak operator form of the Pettis integral. First let us introduce weak operator measurability.

Definition 4.1 (weak measurability, weak integrability). Let $(\Omega, \mathcal{A}, \mu)$ be a finite measure space. Let $f : \Omega \rightarrow S(E)$ be given. The function f is said to be weakly (operator) measurable if and only if

$$\langle f(\omega)x, y \rangle$$

is μ measurable for all $x, y \in E$.

We also say that the function f is weakly integrable if it is weakly measurable and

$$\int_{\Omega} |\langle f(\omega)x, y \rangle| d\mu(\omega) < +\infty$$

for all $x, y \in E$.

Now integrability is defined as follows.

Definition 4.2 (weak operator Pettis integral). Let $(\Omega, \mathcal{A}, \mu)$ be a finite measure space. Let $f : \Omega \rightarrow S(E)$ be weakly operator measurable. If there exists $A \in S(E)$ such that $\langle Ax, y \rangle = \int_{\Omega} \langle f(\omega)x, y \rangle d\mu(\omega)$ for all $x, y \in E$, then we define

$$\int_{\Omega} f d\mu := A.$$

Clearly the uniqueness of the integral is satisfied. The above definition is based on the definition of the Pettis integral in an arbitrary Banach space B , where we require the existence of $x \in B$, for an $f : \Omega \mapsto B$ such that $\langle x, y^* \rangle = \int_{\Omega} \langle f(\omega), y^* \rangle d\mu(\omega)$ where $y^* \in B^*$. In other words first we require the weak measurability of f i.e., $\langle f(\omega), y^* \rangle$ is measurable for all $y^* \in B^*$, and then integrability i.e., the existence of such $x \in B$. Now there is a

different notion of integrability due to Dunford which asks for the existence of $x^{**} \in B^{**}$ such that $\langle x^{**}, y^* \rangle = \int_{\Omega} \langle f(\omega), y^* \rangle d\mu(\omega)$ for all $y^* \in B^*$. It is known that every weakly measurable function f that is weakly integrable (i.e., $\int_{\Omega} |\langle f(\omega), y^* \rangle| d\mu(\omega) < +\infty$ for all $y^* \in B^*$) is Dunford integrable, see Theorem 11.55 in [2]. In order to ensure the existence of the operator $A \in S(E)$ in Definition 4.2 we need some preliminary results.

Lemma 4.1. *Let $(\Omega, \mathcal{A}, \mu)$ be a finite measure space and let $f : \Omega \rightarrow S(E)$ be weakly measurable and weakly integrable. Then the weak operator Pettis integral $\int_{\Omega} f d\mu$ exists.*

Proof. By assumption $\langle f(\omega)x, y \rangle$ is μ -measurable and $\int_{\Omega} |\langle f(\omega)x, y \rangle| d\mu(\omega) < +\infty$ for all $x, y \in E$. Define $F_x : \Omega \rightarrow E$ by $F_x(\omega) := f(\omega)x$. Then we have for fixed $x \in E$ that for all $y \in E$, $\langle F_x(\omega), y \rangle$ is μ -measurable and $\int_{\Omega} |\langle F_x(\omega), y \rangle| d\mu(\omega) < +\infty$, and since E is Hilbert therefore the linear functionals $y^*(\cdot) := \langle \cdot, y \rangle$ exhaust E^* by the Riesz representation theorem. Therefore the function $F_x : \Omega \rightarrow E$ is Dunford integrable i.e., there exists a unique $\hat{F}(x) \in E^{**} \simeq E$ such that $\int_{\Omega} \langle F_x(\omega), y \rangle d\mu(\omega) = \langle \hat{F}(x), y \rangle$. Now it is routine to check that $x \mapsto \hat{F}(x)$ defines a bounded (it maps bounded sets to bounded sets) linear operator on E , moreover by the assumptions $\hat{F} \in S(E)$. \square

We have a generalized version of the Dominated Convergence Theorem for the Pettis integral that makes us of:

Theorem 4.2. *Let $f : \Omega \mapsto S(E)$ satisfy the following:*

- *There exists a sequence of weak operator Pettis integrable functions f_n such that $\lim_{n \rightarrow \infty} \langle f_n x, y \rangle = \langle f x, y \rangle$ in measure for all $x, y \in E$.*
- *There exists a real valued μ -integrable function $h : \Omega \rightarrow \mathbb{R}$ such that for each linear functional $l(Z) := \langle Zx, y \rangle$ in the weak*-compact polar of some neighborhood generated by the duality provided by the weak operator topology and $n \in \mathbb{N}$ the inequality $|\langle f_n x, y \rangle| \leq h$ holds a.e.*

Then f is weak operator Pettis integrable and

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu$$

in the weak operator topology.

This theorem is a consequence of the Lebesgue Dominated Convergence Theorem and the Banach space form of it can be found in [30].

It is well known that all Bochner integrable functions with respect to a measure are Pettis integrable. It is also known that a Bochner integrable function always has separable range. That is one reason why we cannot apply the Bochner integral, since $S(E)$ contains non-separable subspaces. Also if Ω is a compact Hausdorff space with a probability measure μ and $f : \Omega \mapsto X$ is continuous where X is a Banach space then f is Pettis integrable with respect to μ , see Theorem 3.27 in [36].

On \mathbb{P} the partial order induces a complete metric space structure [40]. The Thompson or part metric is defined as

$$(28) \quad d_\infty(A, B) = \max \{ \log M(A/B), M(B/A) \}$$

for any $A, B \in \mathbb{P}$, where $M(A/B) = \inf \{ \alpha : A \leq \alpha B \}$. The metric space (\mathbb{P}, d_∞) is complete and has some several other nice properties.

Lemma 4.3 (Lemma 10.1 in [23]). *We have*

- (1) $d_\infty(rA, rB) = d_\infty(A, B)$ for any $r > 0$,
- (2) $d_\infty(A^{-1}, B^{-1}) = d_\infty(A, B)$,
- (3) $d_\infty(MAM^*, MBM^*) = d_\infty(A, B)$ for all $M \in \text{GL}(E)$ where $\text{GL}(E)$ denotes the Banach-Lie group of all invertible bounded linear operators on E ,
- (4) $d_\infty(\sum_{i=1}^k t_i A_i, \sum_{i=1}^k t_i B_i) \leq \max_{1 \leq i \leq k} d_\infty(A_i, B_i)$ where $t_i > 0$,
- (5) $e^{-d_\infty(A, B)} B \leq A \leq e^{d_\infty(A, B)} B$ and $e^{-d_\infty(A, B)} A \leq B \leq e^{d_\infty(A, B)} A$.

Property 4. in Lemma 4.3 is important for us, but we need a refined and also a weighted version of it. The refined version considers integrals instead of sums.

Lemma 4.4. *Let $(\Omega, \mathcal{A}, \mu)$ be a probability space, and let $f_1, f_2 : \Omega \mapsto \mathbb{P}$ be measurable such that both f_1 and f_2 are weak operator Pettis integrable with respect to μ . Then*

$$(4') \quad d_\infty(\int_\Omega f_1 d\mu, \int_\Omega f_2 d\mu) \leq \sup_{\omega \in \text{supp}(\mu)} d_\infty(f_1(\omega), f_2(\omega)), \text{ assuming that the supremum is finite.}$$

Proof. Let $\alpha := e^{\sup_{\omega \in \text{supp}(\mu)} d_\infty(f_1(\omega), f_2(\omega))}$. Then for all $\omega \in \text{supp}(\mu)$ we have

$$\begin{aligned} f_1(\omega) &\leq \alpha f_2(\omega) \\ f_2(\omega) &\leq \alpha f_1(\omega). \end{aligned}$$

Now $f_1(\omega) \leq \alpha f_2(\omega)$ is equivalent to

$$\langle [\alpha f_2(\omega) - f_1(\omega)]x, x \rangle \geq 0$$

for all $x \in E$. Then by the definition of the weak operator Pettis integral we have

$$\begin{aligned} \int_\Omega \langle [\alpha f_2(\omega) - f_1(\omega)]x, x \rangle d\mu &\geq 0 \\ \left\langle \left\{ \int_\Omega [\alpha f_2(\omega) - f_1(\omega)] d\mu \right\} x, x \right\rangle &\geq 0 \\ \left\langle \left[\alpha \int_\Omega f_2(\omega) d\mu - \int_\Omega f_1(\omega) d\mu \right] x, x \right\rangle &\geq 0 \\ \int_\Omega f_1 d\mu &\leq \alpha \int_\Omega f_2 d\mu. \end{aligned}$$

Now similar computation leads to $\int_\Omega f_2 d\mu \leq \alpha \int_\Omega f_1 d\mu$ starting from the other inequality. \square

Here is the weighted version of property (4) for finite sums.

Proposition 4.5. *Let $c_i > 0$ be real numbers, $A_i, B_i \in \mathbb{P}$, $1 \leq i \leq 2$ and suppose that $d_\infty(A_1, B_1) \geq d_\infty(A_2, B_2)$. Then we have*

$$e^{d_\infty(c_1 A_1 + c_2 A_2, c_1 B_1 + c_2 B_2)} \leq \max \left\{ \frac{c_1 e^{d_\infty(A_1, B_1)} + c_2 e^{-d_\infty(A_1, A_2)} e^{d_\infty(A_2, B_2)}}{c_1 + c_2 e^{-d_\infty(A_1, A_2)}}, \frac{c_1 e^{d_\infty(A_1, B_1)} + c_2 e^{-d_\infty(B_1, B_2)} e^{d_\infty(A_2, B_2)}}{c_1 + c_2 e^{-d_\infty(B_1, B_2)}} \right\}.$$

Proof. Let $\alpha_i = e^{d_\infty(A_i, B_i)}$. By definition of d_∞ , in order to calculate $d_\infty\left(\sum_{i=1}^2 c_i A_i, \sum_{i=1}^2 c_i B_i\right)$, we are looking for the infimum of all $\beta \geq 0$ such that both

$$(29) \quad \begin{aligned} c_1 A_1 + c_2 A_2 &\leq \beta(c_1 B_1 + c_2 B_2), \\ c_1 B_1 + c_2 B_2 &\leq \beta(c_1 A_1 + c_2 A_2) \end{aligned}$$

are satisfied. We also have that

$$\begin{aligned} c_1 A_1 + c_2 A_2 &\leq c_1 \alpha_1 B_1 + c_2 \alpha_2 B_2 \\ c_1 B_1 + c_2 B_2 &\leq c_1 \alpha_1 A_1 + c_2 \alpha_2 A_2. \end{aligned}$$

From this it follows that the infimum of the β satisfying (29) are bounded above by the infimum of $\beta \geq 0$ satisfying

$$(30) \quad \begin{aligned} c_1 A_1 + c_2 A_2 &\leq c_1 \alpha_1 B_1 + c_2 \alpha_2 B_2 \leq \beta(c_1 B_1 + c_2 B_2), \\ c_1 B_1 + c_2 B_2 &\leq c_1 \alpha_1 A_1 + c_2 \alpha_2 A_2 \leq \beta(c_1 A_1 + c_2 A_2). \end{aligned}$$

The first inequality above is equivalent to

$$0 \leq \sum_{i=1}^2 (\beta - \alpha_i) c_i B_i.$$

Now by Property 4. in Lemma 4.3 we have the natural bound $\beta \leq \max_{1 \leq i \leq 2} \alpha_i$. So we may try to find a better bound by assuming that $\alpha_1 \geq \beta \geq \alpha_2$, where without loss of generality $\alpha_1 \geq \alpha_2$. Using the assumption on β we get that we seek the infimum of all $\beta \geq 0$ such that

$$(\beta - \alpha_2) c_2 B_2 \geq (\alpha_1 - \beta) c_1 B_1.$$

Using property 5. in Lemma 4.3 we have that the sought β above is bounded above by the infimum of all $\beta \geq 0$ satisfying

$$(\beta - \alpha_2) c_2 e^{-d_\infty(B_1, B_2)} \geq (\alpha_1 - \beta) c_1.$$

This is equivalent to

$$\sum_{i=1}^2 (\beta - \alpha_i) c_i e^{-d_\infty(B_1, B_i)} \geq 0,$$

in other words we have that

$$\beta \geq \frac{\sum_{i=1}^2 c_i \alpha_i e^{-d_\infty(B_1, B_i)}}{\sum_{i=1}^2 c_i e^{-d_\infty(B_1, B_i)}} = \frac{\sum_{i=1}^2 e^{d_\infty(A_i, B_i)} c_i e^{-d_\infty(B_1, B_i)}}{\sum_{i=1}^2 c_i e^{-d_\infty(B_1, B_i)}}.$$

Performing the same calculation (A_i in place of B_i) by starting with the second inequality in (30) we get

$$\beta \geq \frac{\sum_{i=1}^2 e^{d_\infty(A_i, B_i)} c_i e^{-d_\infty(A_1, A_i)}}{\sum_{i=1}^2 c_i e^{-d_\infty(A_1, A_i)}}.$$

This means that

$$e^{d_\infty(\sum_{i=1}^2 c_i A_i, \sum_{i=1}^2 c_i B_i)} \leq \max \left\{ \frac{\sum_{i=1}^2 e^{d_\infty(A_i, B_i)} c_i e^{-d_\infty(A_1, A_i)}}{\sum_{i=1}^2 c_i e^{-d_\infty(A_1, A_i)}}, \frac{\sum_{i=1}^2 e^{d_\infty(A_i, B_i)} c_i e^{-d_\infty(B_1, B_i)}}{\sum_{i=1}^2 c_i e^{-d_\infty(B_1, B_i)}} \right\}.$$

which is what we wanted to prove. \square

Let $\overline{B}_A(r) = \{X \in \mathbb{P} : d_\infty(A, X) \leq r\}$. The following elementary inequality for the weighted arithmetic mean will be useful. Let Δ_n denote the convex set of positive probability n-vectors i.e., $\omega = (w_1, \dots, w_n) \in \Delta_n$ if $w_i > 0$ and $\sum_{i=1}^n w_i = 1$.

Lemma 4.6. *Let $a_i \geq 0$, $1 \leq i \leq n$, and suppose that $a_1 \geq a_i$, for all $1 \leq i \leq n$. Let $\omega = (w_1, \dots, w_n) \in \Delta_n$ and let $\mu = (u_1, \dots, u_n) \in \Delta_n$ such that $w_1 \geq u_1$, and $w_i \leq u_i$ for $2 \leq i \leq n$. Then we have*

$$(*) \quad \sum_{i=1}^n w_i a_i \geq \sum_{i=1}^n u_i a_i.$$

Proof. We have

$$\begin{aligned} \sum_{i=1}^n w_i a_i - \sum_{i=1}^n u_i a_i &= (w_1 - u_1) a_1 - \sum_{i=2}^n (u_i - w_i) a_i \\ &\geq (w_1 - u_1) a_1 - \sum_{i=2}^n (u_i - w_i) a_1 \\ &= \sum_{i=1}^n w_i a_1 - \sum_{i=1}^n u_i a_1 = 0. \end{aligned}$$

\square

A cone C is *almost Archimedean* if the closure of the intersection of the cone with any two-dimensional subspace is still a cone, that is, contains no nontrivial subspace. Note that \mathbb{P} is almost Archimedean, so we can use the following contraction result for the vector addition in \mathbb{P} .

We will need a basic contraction result for the weighted arithmetic and harmonic means on the cone \mathbb{P} . The following contraction result has already appeared in [27].

Theorem 4.7 (Theorem 2.6 [27]). *Let C be an almost Archimedean cone in a real vector space V such that C does not contain 0 and consists of only one part. Let $a, b \in C$. Then with respect to the Thompson metric the translation $\tau_a(x) := a + x$ restricted to $\{x \in C : x \leq b\}$ has Lipschitz constant $\frac{M(b/a)}{M(b/a)+1}$.*

Since \mathbb{P} is almost Archimedean the above is applicable, however we state here our own contraction result below. Even though our Lipschitz constant is not as sharp as in the above, its proof will serve as a reference point for more involved calculations to come.

Lemma 4.8. *Let $a, b > 0$ be real numbers. Then the mappings $h_{a,b,A}^+(B) = aA + bB$ and $h_{a,b,A}^-(B) = (aA^{-1} + bB^{-1})^{-1}$ are strict contractions on every $\overline{B}_A(r)$ for all $r < \infty$ i.e., for all $X, Y \in \overline{B}_A(r)$*

$$d_\infty(h_{a,b,A}^\pm(X), h_{a,b,A}^\pm(Y)) \leq \rho d_\infty(X, Y)$$

where

$$\rho = \frac{\log \frac{be^{3r}+a}{be^r+a}}{2r}$$

for a fixed $0 < r < \infty$.

Proof. It suffices to prove the above for $h_{a,b,A}^+(B)$, since then the same follows for $h_{a,b,A}^-(B)$ by the inversion invariance of the metric d_∞ : property (2) in Lemma 4.3. Also by property (3) in Lemma 4.3 it is enough to prove for the case when $A = I$. Let $X, Y \in \overline{B}_I(r)$. By Proposition 4.5 we have that

$$e^{d_\infty(h_{a,b,I}^+(X), h_{a,b,I}^+(Y))} \leq \max \left\{ \frac{ae^{-d_\infty(X,I)} + be^{d_\infty(X,Y)}}{ae^{-d_\infty(X,I)} + b}, \frac{ae^{-d_\infty(Y,I)} + be^{d_\infty(X,Y)}}{ae^{-d_\infty(Y,I)} + b} \right\}$$

Since $X, Y \in \overline{B}_I(r)$ we have that

$$e^{-r}I \leq X, Y \leq e^rI$$

which means that

$$d_\infty(X, I), d_\infty(Y, I) \leq r.$$

Hence it follows that

$$\frac{ae^{-d_\infty(Y,I)}}{ae^{-d_\infty(Y,I)} + b} = \frac{1}{1 + e^{d_\infty(Y,I)}b/a} \geq \frac{1}{1 + e^rb/a} = \frac{ae^{-r}}{ae^{-r} + b}$$

and similarly

$$\frac{ae^{-d_\infty(X,I)}}{ae^{-d_\infty(X,I)} + b} \geq \frac{ae^{-r}}{ae^{-r} + b}.$$

Also

$$\frac{b}{ae^{-d_\infty(X,I)} + b} \leq \frac{b}{ae^{-r} + b}$$

$$\frac{b}{ae^{-d_\infty(Y,I)} + b} \leq \frac{b}{ae^{-r} + b},$$

hence using (*) this means that

$$(31) \quad e^{d_\infty(h_{a,b,I}^+(X), h_{a,b,I}^+(Y))} \leq \frac{ae^{-r} + be^{d_\infty(X,Y)}}{ae^{-r} + b}.$$

So we seek $0 < \rho < 1$ such that

$$d_\infty(h_{a,b,I}^+(X), h_{a,b,I}^+(Y)) \leq \log \left(\frac{ae^{-r} + be^{d_\infty(X,Y)}}{ae^{-r} + b} \right) \leq \rho d_\infty(X, Y)$$

for all $X, Y \in \overline{B}_I(r)$ i.e., $d_\infty(X, Y) \leq 2r$. It therefore suffices to find the maximum of the function

$$f(x) = \frac{\log \left(\frac{e^x be^r + a}{be^r + a} \right)}{x}$$

on the interval $[0, 2r]$. First routine calculations show that

$$\lim_{x \rightarrow 0+} \frac{\log \left(\frac{e^x be^r + a}{be^r + a} \right)}{x} = \frac{be^r}{be^r + a} < 1.$$

Then the maximization problem is the same as finding the smallest $\rho < 1$ such that

$$\log \left(\frac{xbe^r + a}{be^r + a} \right) \leq \rho \log(x)$$

on the transformed interval $[1, e^{2r}]$. But that is equivalent to

$$\frac{xbe^r}{be^r + a} + \frac{a}{be^r + a} \leq x^\rho.$$

Since for $0 < \rho < 1$ the function x^ρ is concave monotonically increasing, $1^\rho = 1$ and its derivative is ρ at 1 therefore the smallest such ρ is determined at the e^{2r} endpoint of the interval, hence

$$\rho = \frac{\log \left(\frac{e^{2r} be^r + a}{be^r + a} \right)}{2r}.$$

□

Remark 4.1. By Proposition 4.5 it is clear that the functions $h_{a,b,A}^\pm(X)$ for all $a, b \geq 0$ are nonexpansive on the whole \mathbb{P} i.e.,

$$d_\infty(h_{a,b,A}^\pm(X), h_{a,b,A}^\pm(Y)) \leq d_\infty(X, Y).$$

Remark 4.2. By Lemma 4.8 it follows that the weighted arithmetic $f(B) := (1-s)A + sB$ and harmonic $g(B) := ((1-s)A^{-1} + sB^{-1})^{-1}$ means are strict contractions on $\overline{B}_A(r)$ for all $r < \infty$ and $s \in [0, 1]$. The contraction coefficients ρ are strictly less than 1 for all $s \in (0, 1)$, in general the ρ calculated explicitly in the proof of Lemma 4.8 is a monotonically decreasing function in s and $\rho \rightarrow 1$ as $s \rightarrow 1-$ also $\rho \rightarrow 0$ as $s \rightarrow 0+$. The case $s = 1$ gives the right trivial mean $M(A, B) = B$ that is nonexpansive i.e.,

$$d_\infty(M(A, X), M(A, Y)) \leq d_\infty(X, Y),$$

while $s = 0$ is the left trivial mean $M(A, B) = A$ and it has contraction coefficient 0 on all of \mathbb{P} .

These preliminary results yield the following contraction result.

Theorem 4.9. *Let $M \in \mathfrak{M}$ and $f(X) = M(A, X)$. If M is not the right trivial mean (i.e., $M(A, B) \neq B$) then the mapping $f(X)$ is a strict contraction on $\overline{B}_A(r)$ for all $r < \infty$ i.e., there exists $0 < \rho_r < 1$ such that*

$$d_\infty(f(X), f(Y)) \leq \rho_r d_\infty(X, Y)$$

for all $X, Y \in \overline{B}_A(r)$.

If M is the right trivial mean (i.e., $M(A, B) = B$) then $f(X)$ is nonexpansive on \mathbb{P} , that is

$$d_\infty(f(X), f(Y)) \leq d_\infty(X, Y)$$

for all $A, X, Y \in \mathbb{P}$.

Proof. The case of the right trivial mean is just the preceding remark, so assume that $M(A, B)$ is not the right trivial mean. Again by property 3 in Lemma 4.3 it is enough to prove for the case when $A = I$. By Proposition 3.2 the mean $M \in \mathfrak{M}$ is represented as

$$(32) \quad M(I, X) = \int_{[0,1]} [(1-s)I + sX^{-1}]^{-1} d\nu(s) = \int_{[0,1]} h_s(X) d\nu(s)$$

So let $X, Y \in \overline{B}_I(r)$. There are other simple cases when the probability measure is supported only over the two points $\{0\}, \{1\}$. These cases include the weighted arithmetic mean with $s \in (0, 1)$ and the case of the left trivial mean which is covered in the preceding remark and are clearly strict contractions on $\overline{B}_I(r)$.

For the remaining cases we split the integral in (32) to the sum of integrals over the mutually disjoint intervals $I_1 = [0, a]$, $I_2 = (a, 1]$ for some $a \in (0, 1)$ such that ν has nonzero mass on the interval I_1 . Such an a clearly exists since we have excluded the case when the measure ν is supported only on the point $\{1\}$. We have that

$$f_i(X) = \int_{I_i} [(1-s)I + sX^{-1}]^{-1} d\nu(s).$$

Notice that the integrals with respect to the measures defining f_i are finite positive measures, so with appropriate rescaling we can consider the integration with respect to a probability measure times a constant. Also the interval $[0, 1]$ is a compact Hausdorff space and the function $h_s(X)$ mapping to \mathbb{P} is continuous with respect to $s \in [0, 1]$, so Theorem 3.27 in [36] applies and $h_s(X)$ is integrable and therefore property (4') also applies for the metric d_∞ . So the functions f_1 and f_2 are nonexpansive due to the preceding remarks i.e., we have

$$d_\infty(f_i(X), f_i(Y)) \leq \rho_i d_\infty(X, Y)$$

where $\rho_2 = 1$. Again due to property (4'), the preceding remarks and Lemma 4.8 we have that

$$\rho_1 = \frac{\log \left(\frac{e^{2r} b e^r + a}{b e^r + a} \right)}{2r}$$

since it is easy to see that by Lemma 4.8 the contraction coefficient ρ calculated in Lemma 4.8 corresponding to a $h_s(X)$ with $s \in [0, a]$ is bounded from above by the contraction coefficient ρ calculated in Lemma 4.8 corresponding to $h_a(X)$.

Now by Proposition 4.5 we have

$$\begin{aligned} e^{d_\infty(f(X), f(Y))} &\leq \max \left\{ \frac{\sum_{i=1}^2 e^{d_\infty(f_i(X), f_i(Y))} e^{-d_\infty(f_m(X), f_i(X))}}{\sum_{i=1}^2 e^{-d_\infty(f_m(X), f_i(X))}}, \right. \\ &\quad \left. \frac{\sum_{i=1}^2 e^{d_\infty(f_i(X), f_i(Y))} e^{-d_\infty(f_m(Y), f_i(Y))}}{\sum_{i=1}^2 e^{-d_\infty(f_m(Y), f_i(Y))}} \right\} \\ &\leq \max \left\{ \frac{\sum_{i=1}^2 e^{\rho_i d_\infty(X, Y)} e^{-d_\infty(f_m(X), f_i(X))}}{\sum_{i=1}^2 e^{-d_\infty(f_m(X), f_i(X))}}, \right. \\ &\quad \left. \frac{\sum_{i=1}^2 e^{\rho_i d_\infty(X, Y)} e^{-d_\infty(f_m(Y), f_i(Y))}}{\sum_{i=1}^2 e^{-d_\infty(f_m(Y), f_i(Y))}} \right\} \end{aligned}$$

where m is such that $d_\infty(f_m(X), f_m(Y)) \geq d_\infty(f_i(X), f_i(Y))$, $i = 1, 2$. To obtain the second inequality above we used the monotonicity of the functions e^x and the weighted arithmetic mean with weights of the form $w_i = e^{-d_\infty(f_m(Y), f_i(Y))}$. The next step is to see that $d_\infty(f_j(Y), f_i(Y))$ is bounded for $i, j = 1, 2$. We have that $Y \in \overline{B}_I(r)$, hence we have

$$e^{-r} I \leq Y \leq e^r I$$

and by the monotonicity of the harmonic mean this yields

$$e^{-r} I \leq [(1-s)I + sY^{-1}]^{-1} \leq e^r I.$$

Integrating this we get

$$\int_{I_i} d\nu(s) e^{-r} I \leq f_i(Y) \leq \int_{I_i} d\nu(s) e^r I$$

for $i = 1, 2$. From this it follows that

$$\max_{i,j=1,2} \sup_{X \in \overline{B}_I(r)} d_\infty(f_j(X), f_i(X)) \leq L$$

where

$$L = 2r + \left| \log \left(\int_{I_1} d\nu(s) \right) - \log \left(\int_{I_2} d\nu(s) \right) \right|$$

and clearly $L < \infty$ since $\int_{I_i} d\nu(s) > 0$. Now since $\rho_2 = 1$ we have that

$$\begin{aligned} e^{d_\infty(f(X), f(Y))} &\leq \max \left\{ \frac{\sum_{i=1}^2 e^{\rho_i d_\infty(X, Y)} e^{-d_\infty(f_m(X), f_i(X))}}{\sum_{i=1}^2 e^{-d_\infty(f_m(X), f_i(X))}}, \right. \\ &\quad \left. \frac{\sum_{i=1}^2 e^{\rho_i d_\infty(X, Y)} e^{-d_\infty(f_m(Y), f_i(Y))}}{\sum_{i=1}^2 e^{-d_\infty(f_m(Y), f_i(Y))}} \right\} \\ &\stackrel{(*)}{\leq} \frac{e^{d_\infty(X, Y)} + e^{\rho_1 d_\infty(X, Y)} e^{-L}}{e^{-L} + 1}, \end{aligned}$$

where to get the last inequality we used a similar argument as we did to obtain (31). Similarly as in the end of the proof of Lemma 4.8 we seek some $0 < \rho < 1$ such that

$$\log \left(\frac{e^{d_\infty(X, Y)} + e^{\rho_1 d_\infty(X, Y)} e^{-L}}{e^{-L} + 1} \right) \leq \rho d_\infty(X, Y)$$

for all $X, Y \in \overline{B}_I(r)$ i.e., $d_\infty(X, Y) \leq 2r$. By the same argument as in the end of the proof of Lemma 4.8 we see that

$$\rho = \frac{\log \left(\frac{e^{2r} + e^{\rho_1 2r} e^{-L}}{e^{-L} + 1} \right)}{2r}$$

suffices and clearly $\rho < 1$. □

Remark 4.3. In [23] Lawson and Lim provided an extension of the geometric, logarithmic and some other iterated means to several variables over \mathbb{P} relying on the Ando-Li-Mathias construction provided in [3]. They established the above contractive property for these means. Our Theorem 4.9 shows that in fact the construction is applicable to all operator means due to the contraction result Theorem 4.9, hence providing multivariable extensions which work in the possibly infinite dimensional setting of \mathbb{P} . This was only known in the finite dimensional setting so far which case was proved in [34].

The further importance of Theorem 4.9 will be apparent in the following sections, when we consider matrix (in fact operator) equations similarly to the case of the matrix power means. We close the section with a general nonexpansive property.

Proposition 4.10. *Let $M : \mathbb{P}^k \rightarrow \mathbb{P}$ be such that*

- (1) if $A_i \leq B_i$ for all $1 \leq i \leq k$, then $M(A_1, \dots, A_k) \leq M(B_1, \dots, B_k)$,
- (2) if $t > 0$, then $M(tA_1, \dots, tA_k) = tM(A_1, \dots, A_k)$,

then

$$d_\infty(M(A_1, \dots, A_k), M(B_1, \dots, B_k)) \leq \max_{1 \leq i \leq k} d_\infty(A_i, B_i)$$

for all $A_i, B_i \in \mathbb{P}$.

Proof. Let $t = \max_{1 \leq i \leq k} d_\infty(A_i, B_i)$. Then $A_i \leq tB_i$ and $B_i \leq tA_i$ for all $1 \leq i \leq k$, so by property 1 and 2

$$\begin{aligned} M(A_1, \dots, A_k) &\leq M(tB_1, \dots, tB_k) = tM(B_1, \dots, B_k) \\ M(B_1, \dots, B_k) &\leq M(tA_1, \dots, tA_k) = tM(A_1, \dots, A_k) \end{aligned}$$

i.e.,

$$\begin{aligned} M(A_1, \dots, A_k) &\leq \max_{1 \leq i \leq k} d_\infty(A_i, B_i) M(B_1, \dots, B_k) \\ M(B_1, \dots, B_k) &\leq \max_{1 \leq i \leq k} d_\infty(A_i, B_i) M(A_1, \dots, A_k). \end{aligned}$$

□

5. GENERALIZED OPERATOR MEANS VIA CONTRACTION PRINCIPLE

In [28] Lim and the author defined the one parameter family of matrix power means $P_s(\omega; \mathbb{A})$ as the unique positive definite solution of the equation

$$(33) \quad X = \sum_{i=1}^k w_i X \#_s A_i$$

where $s \in [-1, 1]$, $w_i > 0$, $\sum_{i=1}^k w_i = 1$ and $A_i \in \mathbb{P}$ and

$$A \#_s B = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^s A^{1/2}$$

is again the weighted geometric mean. Existence and uniqueness of the solution of (33) follow from the fact that the function

$$f(X) = \sum_{i=1}^k w_i X \#_s A_i$$

is a strict contraction for $s \in [-1, 1]$, $s \neq 0$ with respect to Thompson's part metric [28].

Consider the following one-parameter family of real functions:

$$(34) \quad f_{s,t}(x) = \frac{[(1-t)(1-s)+t]x + s(1-t)}{(1-t)(1-s)x + t + s(1-t)}$$

for $t, s \in [0, 1]$. By simple calculation one finds

$$\begin{aligned} (35) \quad f_{s,t}(x) &= \frac{1}{t + s(1-t)} \left\{ s(1-t) + \frac{t}{(1-t)(1-s) + [t + s(1-t)]x^{-1}} \right\} \\ &= \frac{1}{t + s(1-t)} [s(1-t) + th_{t+s(1-t)}(x)]. \end{aligned}$$

Similarly we have for the inverse function $f_{s,t}^{-1}$ that

$$(36) \quad f_{s,t}^{-1}(x) = \frac{[1 - (1-t)s]x - (1-t)}{-(1-t)s(1-s)x + s + (1-s)t}.$$

Lemma 5.1. *The real functions $f_{s,t}(x)$ for any fixed $t, s \in [0, 1]$ defined by (34) are operator monotone and positive (as functions of x) for all $x \in (0, \infty)$.*

Proof. Positivity of $f_{s,t}$ is clear, operator monotonicity follows from the fact that the functions

$$\frac{ax + b}{cx + d}$$

are operator monotone for real $x \neq -\frac{d}{c}$ if $ad - bc > 0$, see for example [8].

Alternatively one can directly arrive at the conclusion for $f_{s,t}$ by looking at (35), so that $f_{s,t}$ is a convex combination of two operator monotone functions. \square

Due to Lemma 5.1 the functions $f_{s,t} \in \mathfrak{m}$ are representing functions of operator means $M_{s,t}(A, B)$ in \mathfrak{M} :

$$(37) \quad M_{s,t}(A, B) := A^{1/2} f_{s,t} \left(A^{-1/2} B A^{-1/2} \right) A^{1/2}.$$

There is an important representation that holds for $f_{s,t}$.

Proposition 5.2. *For all $s, t \in [0, 1]$ we have*

$$f_{s,t}(x) = l_s^{-1}(t l_s(x)),$$

where $l_s(x) := \frac{x-1}{(1-s)x+s}$ and l_s^{-1} is its inverse function. Moreover $f_{s,t} \in \mathfrak{m}(t)$.

Proof. By simple computation. \square

Remark 5.1. Notice that the functions l_s are the ones that occur in the integral representation given in Corollary 2.1 for the functions in \mathfrak{L} . This basic observation will be of fundamental importance for us.

There are some other basic properties that follow from Proposition 5.2. For example we have a semi-group property for the functions $f_{s,t}$:

$$(38) \quad f_{s,t_1} \circ f_{s,t_2}(x) = f_{s,t_1 t_2}(x).$$

We will make use of the operator mean corresponding to the transpose of $f_{s,t}(x)$ i.e.,

$$\begin{aligned} x f_{s,t}(1/x) &= \frac{1}{t + s(1-t)} \left\{ s(1-t)x + \frac{t}{(1-t)(1-s)x^{-1} + [t + s(1-t)]} \right\} \\ &= \frac{1}{t + s(1-t)} [s(1-t)x + t h_{(1-t)(1-s)}(x)]. \end{aligned}$$

Proposition 5.3. *Let $f(X) := M_{s,t}(X, A)$ with $s \in [0, 1]$, $t \in (0, 1]$ arbitrary. Then the mapping $f(X)$ is a strict contraction on $\overline{B}_A(r)$ for all $r < \infty$ with contraction coefficient*

$$\rho = \rho_1 + \frac{\log \frac{be^{-r} + ae^{2r(1-\rho_1)}}{be^{-r} + a}}{2r},$$

where $\rho_1 = \frac{\log \frac{e^{3r(1-t)} + t}{e^{r(1-t)} + t}}{2r}$ and $a = \frac{s(1-t)}{t+s(1-t)}$, $b = \frac{t}{t+s(1-t)}$.

Proof. Consider the functions

$$\begin{aligned} f_1(X) &:= \{(1-t)(1-s)X^{-1} + [t+s(1-t)]A^{-1}\}^{-1} \\ f_2(X) &:= X. \end{aligned}$$

Then $M_{s,t}(X, A) = \frac{1}{t+s(1-t)}[tf_1(X) + s(1-t)f_2(X)]$ i.e., a convex combination. Now use Lemma 4.8 to conclude that f_1 is a strict contraction on $\overline{B}_A(r)$ with contraction coefficient

$$\rho_1 = \frac{\log \frac{e^{3r(1-t)(1-s)} + t + s(1-t)}{e^{r(1-t)(1-s)} + t + s(1-t)}}{2r} \stackrel{(*)}{\leq} \frac{\log \frac{e^{3r(1-t)} + t}{e^{r(1-t)} + t}}{2r}.$$

The other function f_2 is nonexpansive. Now we have

$$\begin{aligned} d_\infty(f_1(X), f_2(X)) &= d_\infty\left(\{(1-t)(1-s)X^{-1} + [t+s(1-t)]A^{-1}\}^{-1}, X\right) \\ &= d_\infty\left((1-t)(1-s)I + [t+s(1-t)]X^{1/2}A^{-1}X^{1/2}, I\right) \\ &\leq d_\infty(X, A), \end{aligned}$$

where the last inequality follows from

$$\begin{aligned} e^{d_\infty((1-t)(1-s)I + [t+s(1-t)]X^{1/2}A^{-1}X^{1/2}, I)} &\leq \\ \max \left\{ \|(1-t)(1-s)I + [t+s(1-t)]X^{1/2}A^{-1}X^{1/2}\|_\infty, \right. \\ \left. \|(1-t)(1-s)I + [t+s(1-t)]X^{-1/2}AX^{-1/2}\|_\infty \right\} \\ &\leq (1-t)(1-s) + [t+s(1-t)]e^{d_\infty(X, A)} \\ &\leq e^{d_\infty(X, A)} \\ &\leq e^r. \end{aligned}$$

This means that $d_\infty(f_1(X), f_2(X)) \leq r$.

Let $X, Y \in \overline{B}_A(r)$. By Proposition 4.5 we have that

$$\begin{aligned} &e^{d_\infty(M_{s,t}(X, A), M_{s,t}(Y, A))} \\ &\leq \max \left\{ \frac{ae^{d_\infty(f_2(X), f_2(Y))} + e^{-d_\infty(f_1(X), f_2(X))}be^{d_\infty(f_1(X), f_1(Y))}}{a + e^{-d_\infty(f_1(X), f_2(X))}b}, \right. \\ &\quad \left. \frac{ae^{d_\infty(f_2(X), f_2(Y))} + e^{-d_\infty(f_1(Y), f_2(Y))}be^{d_\infty(f_1(X), f_1(Y))}}{a + e^{-d_\infty(f_1(Y), f_2(Y))}b} \right\}. \end{aligned}$$

Since f_1 is a strict contraction on $\overline{B}_A(r)$ with contraction coefficient ρ_1 and also $d_\infty(f_1(X), f_2(X)) \leq r$ for all $X \in \overline{B}_A(r)$, hence we have

$$e^{d_\infty(M_{s,t}(X,A), M_{s,t}(Y,A))} \stackrel{(*)}{\leq} \frac{ae^{d_\infty(X,Y)} + e^{-r}be^{\rho_1 d_\infty(X,Y)}}{a + e^{-r}b}$$

using a similar argument as we did to obtain (31). Now we seek $0 < \rho < 1$ such that

$$\begin{aligned} d_\infty(M_{s,t}(X, A), M_{s,t}(Y, A)) &\leq \log \left(\frac{ae^{d_\infty(X,Y)} + e^{-r}be^{\rho_1 d_\infty(X,Y)}}{a + e^{-r}b} \right) \\ &\leq \rho d_\infty(X, Y) \end{aligned}$$

for all $X, Y \in \overline{B}_I(r)$ i.e., $d_\infty(X, Y) \leq 2r$. It therefore suffices to find the maximum of the function

$$f(x) = \frac{\log \left(\frac{e^{\rho_1 x} be^{-r} + ae^x}{be^{-r} + a} \right)}{x}$$

on the interval $[0, 2r]$. First routine calculations show that

$$\lim_{x \rightarrow 0+} \frac{\log \left(\frac{e^{\rho_1 x} be^{-r} + ae^x}{be^{-r} + a} \right)}{x} = \frac{\rho_1 be^{-r} + a}{be^{-r} + a} < 1.$$

Then the maximization problem is the same as finding the smallest $\rho < 1$ such that

$$\log \left(\frac{x^{\rho_1} be^{-r} + ax}{be^{-r} + a} \right) \leq \rho \log(x)$$

on the transformed interval $[1, e^{2r}]$. But that is equivalent to

$$\frac{x^{\rho_1} be^{-r}}{be^{-r} + a} + \frac{ax}{be^{-r} + a} \leq x^\rho,$$

which by substitution with $y := x^{1-\rho_1}$ is equivalent to requiring

$$\frac{be^{-r}}{be^{-r} + a} + \frac{ay}{be^{-r} + a} \leq y^{\frac{\rho-\rho_1}{1-\rho_1}}$$

for $y \in [1, e^{2r(1-\rho_1)}]$. Now similar considerations as in the end of the proof of Lemma 4.8 lead to that the smallest such $\frac{\rho-\rho_1}{1-\rho_1}$ is determined at the $e^{2r(1-\rho_1)}$ endpoint of the interval, hence

$$\frac{\rho - \rho_1}{1 - \rho_1} = \frac{\log \frac{be^{-r} + ae^{2r(1-\rho_1)}}{be^{-r} + a}}{2r(1 - \rho_1)}.$$

In other words

$$\rho = \rho_1 + \frac{\log \frac{be^{-r} + ae^{2r(1-\rho_1)}}{be^{-r} + a}}{2r}$$

which is strictly less than 1. □

Now we would like to consider convex combinations of $M_{s,t}(A, B)$ which amounts to integrating with respect to a measure. We have already introduced the weak operator Pettis integral in Definition 4.2 which was sufficient for our purpose so far. We will consider probability measures, but now we change the point of view and instead we mostly consider the measures directly given in \mathbb{P} , in other words the push forward measures under the injective continuous map from the probability space to \mathbb{P} .

Definition 5.1. Let $\mathcal{P}(\mathbb{P})$ denote the set of all probability measures with bounded support in \mathbb{P} on the σ -algebra generated by the open sets of \mathbb{P} in the norm topology.

Corollary 5.4. *Let $\sigma \in \mathcal{P}(\mathbb{P})$. Then any norm/strong/weak continuous function $f : \text{supp } \sigma \rightarrow \mathbb{P}$ with bounded range is weak operator Pettis integrable with respect to σ .*

Proof. Let $x, y \in E$. Then $\omega \mapsto \langle f(\omega)x, y \rangle$ is a continuous real valued function, hence measurable. The range of f is bounded, so there exists a $K < \infty$ such that

$$f(\omega) \leq KI$$

for all $\omega \in \text{supp } \sigma$. Hence by the Cauchy-Schwarz inequality we have

$$|\langle f(\omega)x, y \rangle| \leq \|f(\omega)x\| \|y\| \leq K\|x\| \|y\|$$

for fixed $x, y \in E$ i.e., the continuous real function $\langle f(\omega)x, y \rangle$ is bounded, hence integrable by the dominated convergence theorem or Theorem 11.8 in [2]. Therefore we can apply Lemma 4.1. \square

The above corollary ensures us that in the remaining parts of the paper all weak operator Pettis integrals exist, since all of our functions will be continuous and bounded. In particular we will consider various integrals of $l_s(X)$ and $M_{s,t}(X, A)$ which have lower and upper bounds due to (25), (17) and (18), moreover they are norm continuous functions in each of their arguments.

Let $\mathcal{P}([0, 1])$ denote the set of all probability measures over the interval $[0, 1]$. $\mathcal{P}([0, 1])$ is a subset of the Banach space of finite signed measures over the interval $[0, 1]$. Also $\mathcal{P}(\mathbb{P})$ is a subset of the Banach space of finite signed measures over \mathbb{P} and in both cases the norm is provided by the total variation.

Notation. $\mathcal{P}([0, 1] \times \mathbb{P})$ denotes the set of all probability measures on $[0, 1] \times \mathbb{P}$ with bounded support. For $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$ we say that μ is supported on $\{s\} \times \{A\}$ for some $s \in [0, 1]$ and $A \in \mathbb{P}$, if $\{s\} \times \{A\} \subseteq \text{supp } \mu$. We also say that μ is supported in $[0, 1] \times \{A\}$, if there exists some non-empty subset $I \subseteq [0, 1]$ such that $I \times \{A\} \subseteq \text{supp } \mu$.

Definition 5.2 (Partial order for measures). Given a probability space $(\Omega, \mathcal{A}, \pi)$, let $g_1 : \Omega \mapsto \mathbb{P}$ and $g_2 : \Omega \mapsto \mathbb{P}$ be two π -measurable maps.

Denote by $\alpha_1 := g_{1*}\pi$ and $\alpha_2 := g_{2*}\pi$ the pushforward measures. Then we denote by

$$\alpha_1 \leq \alpha_2$$

if and only if $g_1(\omega) \leq g_2(\omega)$ for all $\omega \in \Omega$.

Similarly let $([0, 1] \times \Omega, \mathcal{A}, \mu)$ be a probability space. Let $f_1 : [0, 1] \times \Omega \mapsto [0, 1] \times \mathbb{P}$ and $f_2 : [0, 1] \times \Omega \mapsto [0, 1] \times \mathbb{P}$ be two μ -measurable maps and let $\nu_1 := f_{1*}\mu$ and $\nu_2 := f_{2*}\mu$ denote the pushforward measures on $[0, 1] \times \mathbb{P}$. Then we denote by

$$\nu_1 \leq \nu_2$$

if and only if for all fixed $s \in [0, 1]$ and $\omega \in \Omega$ we have $[f_1(s, \omega)]_1 = [f_2(s, \omega)]_1$ and $[f_1(s, \omega)]_2 \leq [f_2(s, \omega)]_2$ with respect to the positive definite order.

Remark 5.2. The above Definition 5.2 of ordering of measures in $\mathcal{P}([0, 1] \times \mathbb{P})$ is a direct generalization of the ordering of k -tuples $\mathbb{A}, \mathbb{B} \in \mathbb{P}^n$. In previous works [24, 28] $\mathbb{A} \leq \mathbb{B}$ was understood element-wise i.e., $A_i \leq B_i$ for all $1 \leq i \leq k$, and only tuples with identical associated element-wise weights were compared. Under this joint order operator monotonicity of (weighted) multivariable means were derived. Let us consider a typical example. Let ξ be a probability measure on $[0, 1]$, let $(\Omega, \mathcal{A}, \mu)$ be a probability space and let $f_1 : \Omega \mapsto \mathbb{P}$ and $f_2 : \Omega \mapsto \mathbb{P}$ be two μ -measurable maps with $f_1(\omega) \leq f_2(\omega)$ for all $\omega \in \Omega$. Then denoting by $\nu_1 := f_{1*}\mu$ and $\nu_2 := f_{2*}\mu$ the pushforward measures, we have for the product measures $\xi \times \nu_1 \leq \xi \times \nu_2$ in the sense of our order.

Remark 5.3. From the cone-theoretic point of view, the order $\mu_1 \leq \mu_2$ in Definition 5.2 should be defined by requiring $\int_{[0, 1] \times \mathbb{P}} f(s, A) d\mu_1(s, A) \leq \int_{[0, 1] \times \mathbb{P}} f(s, A) d\mu_2(s, A)$ for all measurable and integrable functions $f : [0, 1] \times \mathbb{P} \rightarrow \mathbb{P}$. If we had adopted this order, then it is not hard to see that all of our operator means defined in the remaining parts of the paper would have been monotone with respect to this cone-theoretic order as well, which appears to be weaker than our adopted order above i.e., one can compare more pairs of measures using this cone-theoretic partial order. Also the proof of monotonicity would be along the same lines here below.

The set $\mathcal{P}([0, 1] \times \mathbb{P})$ is a subset of a Banach space of all finite signed measures on $[0, 1] \times \mathbb{P}$ with the total variation norm. Now we need a Fubini type of result.

Lemma 5.5. *Let $\nu \in \mathcal{P}([0, 1])$ and $\sigma \in \mathcal{P}(\mathbb{P})$. Then*

$$\begin{aligned} & \int_{\mathbb{P}} \int_{[0, 1]} l_s \left(X^{-1/2} A X^{-1/2} \right) d\nu(s) d\sigma(A) \\ &= \int_{[0, 1] \times \mathbb{P}} l_s \left(X^{-1/2} A X^{-1/2} \right) d(\nu \times \sigma)(s, A) \\ &= \int_{[0, 1]} \int_{\mathbb{P}} l_s \left(X^{-1/2} A X^{-1/2} \right) d\sigma(A) d\nu(s). \end{aligned}$$

Proof. First note that the function $l_s(X^{-1/2}AX^{-1/2})$ is jointly norm continuous in $\mathbb{P} \times [0, 1]$, so therefore the real function $\langle l_s(X^{-1/2}AX^{-1/2})u, v \rangle$ is also continuous on $[0, 1] \times \mathbb{P}$ for any $u, v \in E$, hence it is measurable with respect to $\nu \times \sigma$, also it is bounded since the support of $\nu \times \sigma$ is bounded i.e., $l_s(X^{-1/2}AX^{-1/2})$ is weak operator Pettis integrable. The rest of the assertion follows from Fubini's theorem for the Lebesgue integrable function $\langle l_s(X^{-1/2}AX^{-1/2})u, v \rangle$. \square

In what follows after the above preparations we will study an analogue of (33).

Lemma 5.6. *Let $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$ and $t \in (0, 1]$. Then the function*

$$(39) \quad f(X) = \int_{[0, 1] \times \mathbb{P}} M_{s,t}(X, A) d\mu(s, A)$$

is a strict contraction with respect to the Thompson metric $d_\infty(\cdot, \cdot)$ on every bounded $S \subseteq \mathbb{P}$ such that $\text{supp } \mu \subseteq [0, 1] \times S$.

Proof. Fix $t \in (0, 1]$. Choose a large enough $r < \infty$ such that for each $A \in \mathbb{P}$ for which μ is supported in $[0, 1] \times \{A\}$, we have $S \subseteq \overline{B}_A(r)$. We can do that since $\text{supp } \mu$ is bounded in $[0, 1] \times \mathbb{P}$ by definition, also therefore we have uniformly

$$e^k I \leq A \leq e^K I$$

for all $A \in \mathbb{P}$ such that μ is supported in $[0, 1] \times \{A\}$ and for some constants $-\infty < k \leq K < \infty$. By Proposition 5.3 $g_{s,t}(X) = M_{s,t}(X, A)$ is a strict contraction on $\overline{B}_A(r)$, with contraction coefficient

$$\rho = \rho_1 + \frac{\log \frac{be^{-r} + ae^{2r(1-\rho_1)}}{be^{-r} + a}}{2r},$$

where $\rho_1 = \frac{\log \frac{e^{3r(1-t)} + t}{e^r(1-t) + t}}{2r}$ and $a = \frac{s(1-t)}{t+s(1-t)}$, $b = \frac{t}{t+s(1-t)}$. Since $t \leq b \leq 1$ and $0 \leq a \leq 1-t$ we have by using (*) that

$$\rho \leq \rho_1 + \frac{\log \frac{te^{-r} + (1-t)e^{2r(1-\rho_1)}}{te^{-r} + (1-t)}}{2r}$$

which is strictly less than 1. Hence using property (4') in Lemma 4.4 we have that

$$d_\infty(f(X), f(Y)) \leq \rho d_\infty(X, Y)$$

for all $X, Y \in S$ i.e., $f(X)$ is a strict contraction on S . \square

Proposition 5.7. *Let $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$, $t \in (0, 1]$. Then the equation*

$$(40) \quad X = \int_{[0, 1] \times \mathbb{P}} M_{s,t}(X, A) d\mu(s, A)$$

has a unique positive definite solution in \mathbb{P} .

Proof. By Lemma 5.6, for every large enough bounded subset $S \subseteq \mathbb{P}$ the function $f(X)$ given in (39) is a strict contraction on S . Suppose now that $S := \overline{B}_I(r)$ for any $r < \infty$ such that $\text{supp } \mu \subseteq [0, 1] \times S$. Then we claim that $f(S) \subseteq S$. We have the following bounds given by (25):

$$[(1-t)X^{-1} + tA^{-1}]^{-1} \leq M_{s,t}(X, A) \leq (1-t)X + tA.$$

Since $S = \overline{B}_I(r)$ it follows that for any $X \in S$ we have $e^{-r}I \leq X \leq e^r I$. Since $\text{supp } \mu \subseteq [0, 1] \times S$ we also have $e^{-r}I \leq A \leq e^r I$ for all $A \in \mathbb{P}$ such that μ is supported in $[0, 1] \times \{A\}$. Therefore by property (ii) in Definition 3.1 we have

$$\begin{aligned} e^{-r}I &= [(1-t)(e^{-r}I)^{-1} + t(e^{-r}I)^{-1}]^{-1} \\ &\leq [(1-t)X^{-1} + t(e^{-r}I)^{-1}]^{-1} \\ &\leq \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu(s, A) \end{aligned}$$

for all $X \in S$ and $A \in \mathbb{P}$ such that μ is supported in $[0, 1] \times \{A\}$. A similar argument using the weighted arithmetic mean leads to

$$\int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu(s, A) \leq e^r I.$$

This yields that $e^{-r}I \leq f(X) \leq e^r I$ for all $X \in S = \overline{B}_I(r)$, hence $f(S) \subseteq S$.

Now the iterates of $f^{on}(X)$ stay in S if $X \in S$ and S is closed. Also f is a strict contraction on S , therefore by Banach's fixed point theorem $f(X)$ has a unique fixed point in S , so equation (40) has a unique positive definite solution in S . Since $S = \overline{B}_I(r)$ was an arbitrary, large enough bounded subset of \mathbb{P} such that $\text{supp } \mu \subseteq [0, 1] \times S$, it follows that the same holds on all of \mathbb{P} . \square

Definition 5.3 (Induced Operator Mean). Let $\sigma \in \mathcal{P}(\mathbb{P})$, $t \in (0, 1]$ and ν be a probability measure on $[0, 1]$. We denote by $L_{t,\nu}(\sigma)$ the unique solution $X \in \mathbb{P}$ of the equation

$$(41) \quad X = \int_{\mathbb{P}} \int_{[0,1]} M_{s,t}(X, A) d\nu(s) d\sigma(A).$$

We call $L_{t,\nu}(\sigma)$ the σ -weighted ν -induced operator mean.

Let $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$ and $t \in (0, 1]$. Then we denote by $L_t(\mu)$ the unique solution $X \in \mathbb{P}$ of the equation

$$(42) \quad X = \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu(s, A).$$

We call $L_t(\mu)$ the μ -weighted induced operator mean.

In the above definition in the notations there should be no confusion, since the number of measures as arguments of L should determine which operator mean is intended.

Remark 5.4. Let $f(X)$ be defined by (39). Then by the monotonicity of $M_{s,t}$, f is monotone: $X \leq Y$ implies that $f(X) \leq f(Y)$.

For any $X \in \text{GL}(E)$ and $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$ we will use the notation

$$(X\mu X^*)(s, A) := \mu(s, X^{-1}A(X^*)^{-1}),$$

similarly for $\sigma \in \mathcal{P}(\mathbb{P})$

$$(X\sigma X^*)(A) := \sigma(X^{-1}A(X^*)^{-1}).$$

Integrating with respect to the above two measures is not a problem due to a generalized form of the change of variables formula that holds for the Lebesgue integral and hence trivially for the weak operator Pettis integral, see for example Theorem 2.26 in [7]. Also weak operator Pettis integration with respect to other transformed measures that we will see later is also permitted due to the continuity of the mappings.

Proposition 5.8. *Let $\mu, \mu_1, \mu_2 \in \mathcal{P}([0, 1] \times \mathbb{P})$ and $t \in (0, 1]$. Then*

- (1) $L_t(\mu) = A$ if $\text{supp } \mu \subseteq [0, 1] \times \{A\}$ for an $A \in \mathbb{P}$;
- (2) $L_t(\mu_1) \leq L_t(\mu_2)$ if $\mu_1 \leq \mu_2$;
- (3) $L_t(X\mu X^*) = XL_t(\mu)X^*$ for any $X \in \text{GL}(E)$;
- (4) Suppose $\int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu_1(s, A) \leq \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu_2(s, A)$ for $\mu_1, \mu_2 \in \mathcal{P}([0, 1] \times \mathbb{P})$. Then $L_t(\mu_1) \leq L_t(\mu_2)$;
- (5) If $0 < t_1 \leq t_2 \leq 1$ then $L_{t_1}(\mu) \leq L_{t_2}(\mu)$;
- (6) If $d\mu_2(s, A) = d\mu_1(s, g(s, A))$ where g is measurable for fixed s , then $(1 - u)L_t(\mu_1) + uL_t(\mu_2) \leq L_t((1 - u)\mu_1 + u\mu_2)$ for any $u \in [0, 1]$;
- (7) If $d\mu_2(s, A) = d\mu_1(s, g(s, A))$ where g is measurable for all fixed s , then $d_\infty(L_t(\mu_1), L_t(\mu_2)) \leq \sup_{\mu_2 \text{ is supported on } \{s\} \times \{A\}} \{d_\infty(A, g(s, A))\}$;
- (8) $\Phi(L_t(\mu)) \leq L_t(\Phi(\mu))$ for any measurable positive unital linear map Φ , where $\Phi(\mu)(s, A) := \mu(s, \Phi^{-1}(A))$.

Proof. (1) By (42) we have $X = \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu(s, A)$ and using that $M_{s,t}(A, A) = A$ we see that $X = A$ is a, and by uniqueness, the solution of (42).

(2) Define

$$f(X) := \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu_1(s, A)$$

and

$$g(X) := \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu_2(s, A).$$

Then $L_t(\mu_1) = \lim_{l \rightarrow \infty} f^{ol}(X)$ and $L_t(\mu_2) = \lim_{l \rightarrow \infty} g^{ol}(X)$ for any $X \in \mathbb{P}$, by the Banach fixed point theorem. By the monotonicity of $M_{s,t} \in \mathfrak{M}$ we have $M_{s,t}(X, g(s, A)) \leq M_{s,t}(X, A)$ and if we integrate this we get $f(X) \leq g(X)$ for all $X \in \mathbb{P}$, similarly follows that $f(X) \leq f(Y), g(X) \leq g(Y)$ whenever $X \leq Y$. Let $X_0 := L_t(\mu_1)$. Then $f(X_0) \leq g(X_0)$ and $f^{o2}(X_0) =$

$f(f(X_0)) \leq g(f(X_0)) \leq g^{\circ 2}(X_0)$. Inductively, we have $f^{\circ l}(X_0) \leq g^{\circ l}(X_0)$ for all $l \in \mathbb{N}$. Therefore, $L_t(\mu_1) = \lim_{l \rightarrow \infty} f^{\circ l}(X_0) \leq \lim_{l \rightarrow \infty} g^{\circ l}(X_0) = L_t(\mu_2)$.

(3) We have $XM_{s,t}(A, B)X^* = M_{s,t}(XAX^*, XBX^*)$ and applying this to the defining equation (41) the property follows.

(4) Define

$$f(X) := \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu_1(s, A)$$

and

$$g(X) := \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu_2(s, A).$$

Then $L_t(\mu_1) = \lim_{l \rightarrow \infty} f^{\circ l}(X)$ and $L_t(\mu_2) = \lim_{l \rightarrow \infty} g^{\circ l}(X)$ for any $X \in \mathbb{P}$, by the Banach fixed point theorem. By assumption we also have that $f(X) \leq g(X)$. Let $Y := L_t(\mu_1)$. Then we have $Y = f(Y) \leq g(Y)$ and inductively $Y = f^{\circ l}(Y) \leq g^{\circ l}(Y)$, hence $L_t(\mu_1) = Y \leq \lim_{l \rightarrow \infty} g^{\circ l}(Y) = L_t(\mu_2)$.

(5) By Proposition 5.2 for all $s, t \in [0, 1]$ we have

$$f_{s,t}(x) = l_s^{-1}(tl_s(x)),$$

so for $0 < t_1 < t_2 \leq 1$ we have $f_{s,t_1}(x) \leq f_{s,t_2}(x)$, since l_s^{-1} is a monotone convex increasing function, which follows from the operator monotonicity, hence concavity of l_s . Then by Proposition 3.4 we get

$$\int_{[0,1] \times \mathbb{P}} M_{s,t_1}(A, B) d\mu(s, A) \leq \int_{[0,1] \times \mathbb{P}} M_{s,t_2}(A, B) d\mu(s, A).$$

Now we may apply a similar argument as in the proof of property (4) to conclude that $L_{t_1}(\mu) \leq L_{t_2}(\mu)$.

(6) Let $X = L_t(\mu_1)$ and $Y = L_t(\mu_2)$. For $u \in [0, 1]$, we set $Z_u = (1 - u)X + uY$. Let

$$f(Z) = \int_{[0,1] \times \mathbb{P}} M_{s,t}(Z, (1 - u)A + ug(s, A)) d\mu_1(s, A).$$

Then by the joint concavity of two-variable operator means (Theorem 3.5 [20])

$$\begin{aligned} Z_u &= (1 - u)X + uY \\ &= \int_{[0,1] \times \mathbb{P}} [(1 - u)M_{s,t}(X, A) + uM_{s,t}(Y, g(s, A))] d\mu_1(s, A) \\ &\leq \int_{[0,1] \times \mathbb{P}} M_{s,t}((1 - u)X + uY, (1 - u)A + ug(s, A)) d\mu_1(s, A) \\ &= f(Z_u). \end{aligned}$$

Inductively, $Z_u \leq f^{\circ l}(Z_u)$ for all $l \in \mathbb{N}$. Therefore, $(1 - u)L_t(\mu_1) + uL_t(\mu_2) = Z_u \leq L_t((1 - u)\mu_1 + u\mu_2)$.

(7) Follows from a similar argument to the proof of Proposition 4.10 using property (2) and (3).

(8) Note that $\Phi(M_{s,t}(A, B)) \leq M_{s,t}(\Phi(A), \Phi(B))$ for any $A, B > 0$ by Proposition 3.3. Then

$$(43) \quad \begin{aligned} \Phi(L_t(\mu)) &= \int_{[0,1] \times \mathbb{P}} \Phi(M_{s,t}(L_t(\mu), A)) d\mu(s, A) \\ &\leq \int_{[0,1] \times \mathbb{P}} M_{s,t}(\Phi(L_{t,\nu_1}(\sigma_1), A), \Phi(A)) d\mu(s, A). \end{aligned}$$

Define

$$f(X) = \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, \Phi(A)) d\mu(s, A).$$

Then $\lim_{l \rightarrow \infty} f^{ol}(X) = L_t(\Phi(\mu))$ for any $X > 0$. By (43), $f(\Phi(L_t(\mu))) \geq \Phi(L_t(\mu))$. Since f is monotonic, $f^{ol}(\Phi(L_t(\mu))) \geq \Phi(L_t(\mu))$ for all $l \in \mathbb{N}$. Thus

$$L_t(\Phi(\mu)) = \lim_{l \rightarrow \infty} f^{ol}(\Phi(L_t(\mu))) \geq \Phi(L_t(\mu)).$$

□

Corollary 5.9. *Suppose $\nu \in \mathcal{P}([0,1])$ and $\sigma \in \mathcal{P}(\mathbb{P})$ and assume that $\sigma(X) := (1-w)\delta_A(X) + w\delta_B(X)$ with $w \in (0,1)$ where $\delta_A(X)$ denotes the Dirac delta supported on A . Then $L_t(\nu \times \sigma)$ is an operator mean in the two variables (A, B) i.e., $L_t(\nu \times \sigma) \in \mathfrak{M}$.*

Proof. By property (3) in Proposition 5.8 it follows that

$$L_t(\nu \times \sigma) = A^{1/2}g(A^{-1/2}BA^{-1/2})A^{1/2}$$

and property (1) yields that $g(I) = I$. By Lemma 5.6 we have that

$$\lim_{l \rightarrow \infty} f^{ol}(X) = g(A^{-1/2}BA^{-1/2})$$

for all $X \in \mathbb{P}$ where

$$f(X) = \int_{[0,1]} (1-w)M_{s,t}(X, I) + wM_{s,t}(X, A^{-1/2}BA^{-1/2}) d\nu(s).$$

We can choose $X = I$ and then

$$\lim_{l \rightarrow \infty} f^{ol}(I) = g(C)$$

where $C = A^{-1/2}BA^{-1/2}$. Also by simple calculation we have that

$$f(X) = (1-w)Xf_{s,t}(X^{-1}) + wXf_{s,t}(X^{-1}C).$$

By property (2) in Proposition 5.8, g is operator monotone. Moreover $f^{ol}(I)$ is an analytic real map in the single variable C for all l , moreover the net $f^{ol}(I)$ converges uniformly on bounded subsets of \mathbb{P} due to the strict contraction property of $f(X)$. Hence the pointwise limit $\lim_{l \rightarrow \infty} f^{ol}(1)$ for positive real (scalar) C is a continuous real map as well and is identical to g by the properties of the functional calculus of self-adjoint operators, since the net $f^{ol}(I)$ converges in norm for all C (the topology generated by the metric d_∞ agrees with the relative Banach space topology [40]). It is also

easy to see that g is positive on $(0, \infty)$ and $g(1) = 1$, hence g is an operator monotone function in \mathbf{m} . So by Theorem 3.2 in [20] we get that $L_t(\nu \times \sigma) = A^{1/2}g(A^{-1/2}BA^{-1/2})A^{1/2}$ is an operator mean in the sense of Definition 3.1. \square

6. GENERALIZED KARCHER EQUATIONS AND ONE PARAMETER FAMILIES OF OPERATOR MEANS

In this section we generalize the results of [28, 24] which were given for the one parameter family of matrix power means. We will provide solutions of nonlinear operator equations that are given in Definition 2.2, this time considered in the setting of the full (possibly infinite dimensional) cone \mathbb{P} . Let us repeat the definition once more:

Definition 6.1 (Generalized Karcher equation). Let $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$. The generalized Karcher equation is the operator equation

$$\int_{[0,1] \times \mathbb{P}} X^{1/2} l_s(X^{-1/2} A X^{-1/2}) X^{1/2} d\mu(s, A) = 0$$

for $X \in \mathbb{P}$.

Proposition 6.1. *The one parameter family of μ -weighted operator means $L_t(\mu)$ are continuous for $t \in (0, 1]$ on any bounded set $S \subseteq \mathbb{P}$ with respect to the topology generated by d_∞ (the norm topology).*

Proof. The induced operator means $L_t(\mu)$ are fixed points of mappings $f(X)$ given in (39) which are strict contractions on any bounded subset of $S \subseteq \mathbb{P}$ according to Lemma 5.6. Therefore on every bounded set $S \subseteq \mathbb{P}$, $L_t(\mu)$ varies continuously in t with respect to the topology generated by the metric d_∞ due to the continuity of fixed points of pointwisely continuous families of strict contractions [31]. \square

Lemma 6.2. *Let $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$. Then for $t \in (0, 1]$ we have*

$$kI \leq L_t(\mu) \leq \int_{[0,1] \times \mathbb{P}} A d\mu(s, A),$$

where $k > 0$ is such that $kI \leq B$ for any $B \in \mathbb{P}$ such that μ is supported in $[0, 1] \times \{B\}$.

Proof. By Lemma 3.1 and Proposition 5.2 we have that

$$(44) \quad M_{s,t}(X, A) \leq (1-t)X + tA.$$

which yields

$$f(X) := \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\nu(s, A) \leq (1-t)X + t \int_{[0,1] \times \mathbb{P}} A d\mu(s, A).$$

If we define $g(X) := (1-t)X + t \int_{[0,1] \times \mathbb{P}} \text{Ad}\mu(s, A)$, then some simple calculation reveals that for any $X \in \mathbb{P}$ we have

$$\lim_{n \rightarrow \infty} g^{\circ n}(X) = \int_{[0,1] \times \mathbb{P}} \text{Ad}\mu(s, A).$$

From this we have

$$L_t(\mu) = \lim_{n \rightarrow \infty} f^{\circ n}(X) \leq \lim_{n \rightarrow \infty} g^{\circ n}(X) = \int_{[0,1] \times \mathbb{P}} \text{Ad}\mu(s, A).$$

The lower bound follows from the proof of Proposition 5.7 where for any large enough bounded ball $S \subseteq \mathbb{P}$ we have $f(S) \subseteq S$. \square

Let us recall the strong topology on \mathbb{P} . The positive definite partial order \leq is strongly continuous, so if $A_n \rightarrow A$, $B_n \rightarrow B$ and $A_n \leq B_n$ then $A \leq B$. Also if A_n is a monotonically decreasing net in \mathbb{P} with respect to \leq and it is bounded from below, then it converges strongly to the infimum of A_n . Similarly if B_n monotonically increases and is bounded from above, then B_n converges strongly to its supremum [41].

Theorem 6.3. *Let $\mu \in \mathcal{P}([0,1] \times \mathbb{P})$. Then there exists $X_0 \in \mathbb{P}$ such that*

$$\lim_{t \rightarrow 0+} L_t(\mu) = X_0.$$

Proof. By property (5) in Proposition 5.8 we have that $t \mapsto L_t(\mu)$ is a decreasing net bounded from below by kI by Lemma 6.2, hence it is convergent in the strong topology as $t \rightarrow 0+$. \square

Definition 6.2 (Lambda operator means). Let $\Lambda_\nu(\sigma) := \lim_{t \rightarrow 0+} L_{t,\nu}(\sigma)$ and call it the σ -weighted ν -lambda operator mean.

Also let $\mu \in \mathcal{P}([0,1] \times \mathbb{P})$. Then we denote $\Lambda(\mu) := \lim_{t \rightarrow 0+} L_t(\mu)$ and call it the μ -weighted lambda operator mean.

Remark 6.1. If we take the one parameter family of matrix power means $P_t(\omega; \mathbb{A})$, which are weighted by finitely supported probability measures ω on the points \mathbb{A} , then it is known that $\lim_{t \rightarrow 0} P_t(\omega; \mathbb{A})$ is the Karcher mean $\Lambda(\omega; \mathbb{A})$ see [28] for the finite dimensional setting and [24] for the general infinite dimensional case.

Theorem 6.4. *Let $\mu, \mu_1, \mu_2 \in \mathcal{P}([0,1] \times \mathbb{P})$ and $t \in (0,1]$. Then*

- (1) $\Lambda(\mu) = A$ if μ is only supported in $[0,1] \times \{A\}$;
- (2) $\Lambda(\mu_1) \leq \Lambda(\mu_2)$ if $\mu_1 \leq \mu_2$;
- (3) $\Lambda(X\mu X^*) = X\Lambda(\mu)X^*$ for any $X \in \text{GL}(E)$;
- (4) Suppose

$$\int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu_1(s, A) \leq \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu_2(s, A)$$

for all $t \in [0,1]$. Then $\Lambda(\mu_1) \leq \Lambda(\mu_2)$;

- (5) If $d\mu_2(s, A) = d\mu_1(s, g(s, A))$ where g is measurable for fixed s , then $(1-u)\Lambda(\mu_1) + u\Lambda(\mu_2) \leq \Lambda((1-u)\mu_1 + u\mu_2)$ for any $u \in [0,1]$;

- (6) If $d\mu_2(s, A) = d\mu_1(s, g(s, A))$ where g is measurable for all fixed s , then $d_\infty(\Lambda(\mu_1), \Lambda(\mu_2)) \leq \sup_{\mu_2 \text{ is supported on } \{s\} \times \{A\}} \{d_\infty(A, g(s, A))\}$;
- (7) $\Phi(\Lambda(\mu)) \leq \Lambda(\Phi(\mu))$ for any measurable positive unital linear map Φ , where $\Phi(\mu)(s, A) := \mu(s, \Phi^{-1}(A))$.
- (8) $kI \leq \Lambda(\mu) \leq \int_{[0,1] \times \mathbb{P}} A d\mu(s, A)$ where $k > 0$ is such that $kI \leq B$ for any $B \in \mathbb{P}$ such that μ is supported in $[0, 1] \times \{B\}$.

Proof. Each of the properties easily follows from Proposition 5.8 and Lemma 6.2 by taking the limit $t \rightarrow 0+$. \square

Now we turn to the study of the generalized Karcher equation

$$(45) \quad \int_{[0,1] \times \mathbb{P}} l_s(X^{-1/2} A X^{-1/2}) d\mu(s, A) = 0$$

for a $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$. We denote by $K(\mu)$ the set of all solutions X of (45) in \mathbb{P} .

Lemma 6.5. *Operator multiplication is strongly continuous on any bounded set.*

Proof. Let $A_l \rightarrow A, B_l \rightarrow B$ strongly, and $\|A_l\|, \|B_l\| \leq K$. Then

$$\|(A_l B_l - AB)x\| \leq \|A_l(B_l - B)x\| + \|(A_l - A)Bx\| \leq K \|(B_l - B)x\| + \|(A_l - A)Bx\|,$$

so $\|(A_l B_l - AB)x\| \rightarrow 0$ as well. \square

Lemma 6.6. *Let Q be an open or closed subset of \mathbb{R} and let $f : Q \rightarrow \mathbb{R}$ be continuous and bounded. Then f is strong operator continuous on the set $S(E)$ of self adjoint operators with spectrum in Q .*

Proof. Special case of Theorem 3.6 in [17]. \square

The consequence of the above is the following

Lemma 6.7. *The functions*

- (1) x^{-1} ,
- (2) $f_{s,t}(x) = l_s^{-1}(t l_s(x))$ for $0 \leq t \leq 1$,
- (3) the mean $M_{s,t}(A, B)$ for $0 \leq t \leq 1$,

are strongly continuous on the order intervals $[e^{-m}I, e^m I]$ for any $m > 0$.

Lemma 6.8. *Let $V \in S(E)$. Then*

$$(46) \quad \lim_{(t,U) \rightarrow (0,V)} \frac{l_s^{-1}(tU) - I}{t} = V,$$

in the strong operator topology.

Proof. The function l_s^{-1} is the inverse of l_s and simple calculation shows that

$$l_s^{-1}(x) = 1 + \frac{x}{1 - (1-s)x}.$$

Now simple calculation shows that

$$\lim_{(t,U) \rightarrow (0,V)} \frac{l_s^{-1}(tU) - I}{t} = \lim_{(t,U) \rightarrow (0,V)} U[I - (1-s)tU]^{-1} = V.$$

□

Theorem 6.9. *The lambda operator mean $\Lambda(\mu)$ satisfies the generalized Karcher equation*

$$\int_{[0,1] \times \mathbb{P}} X^{1/2} l_s(X^{-1/2} A X^{-1/2}) X^{1/2} d\mu(s, A) = 0.$$

Proof. For $0 < t \leq 1$ let $X_t = L_t(\mu)$ and $X_0 := \Lambda(\mu) = \lim_{t \rightarrow 0+} L_t(\mu)$. By Theorem 6.3 $X_t \rightarrow X_0$ strongly monotonically as $t \rightarrow 0+$ and $kI \leq X_0 \leq X_t \leq \int_{[0,1] \times \mathbb{P}} A d\mu(s, A)$. Now choose $m > 0$ such that $A, X_t, X_0 \in [e^{-m}I, e^m I]$ for all A such that μ is supported in $[0, 1] \times \{A\}$. Then also $X_t \in [e^{-m}I, e^m I]$ for $0 \leq t \leq 1$. The order interval $[e^{-m}I, e^m I]$ is closed under inversion, also $1 \leq x^{1/2} \leq x$ for $x \in [1, \infty)$ and $1 \geq x^{1/2} \geq x$ for $x \in (0, 1)$, so $X_t^{-1/2} A X_t^{-1/2} \in [e^{-m}I, e^m I]$ for all A such that μ is supported in $[0, 1] \times \{A\}$. By the previous lemmas therefore $X_t^{-1/2} A X_t^{-1/2} \rightarrow X_0^{-1/2} A X_0^{-1/2}$ strongly. By the strong continuity of l_s

$$U(A) := l_s(X_t^{-1/2} A X_t^{-1/2}) \rightarrow V(A) := l_s(X_0^{-1/2} A X_0^{-1/2}).$$

By Lemma 6.8 in the strong topology we have

$$(47) \quad \lim_{t \rightarrow 0+} \frac{l_s^{-1}(tU(A)) - I}{t} = V(A) = l_s(X_0^{-1/2} A X_0^{-1/2}).$$

By definition $X_t = \int_{[0,1] \times \mathbb{P}} M_{s,t}(X_t, A) d\mu(s, A)$ which is equivalent to

$$\begin{aligned} I &= \int_{[0,1] \times \mathbb{P}} f_{s,t}(X_t^{-1/2} A X_t^{-1/2}) d\mu(s, A) \\ &= \int_{[0,1] \times \mathbb{P}} l_s^{-1} \left(t l_s(X_t^{-1/2} A X_t^{-1/2}) \right) d\mu(s, A), \end{aligned}$$

that is $0 = \int_{[0,1] \times \mathbb{P}} \frac{f_{s,t}(X_t^{-1/2} A X_t^{-1/2}) - I}{t} d\mu(s, A)$. By (47) we have

$$\begin{aligned} 0 &= \lim_{t \rightarrow 0+} \int_{[0,1] \times \mathbb{P}} \frac{f_{s,t}(X_t^{-1/2} A X_t^{-1/2}) - I}{t} d\mu(s, A) \\ (48) \quad &= \int_{[0,1] \times \mathbb{P}} \lim_{t \rightarrow 0+} \frac{f_{s,t}(X_t^{-1/2} A X_t^{-1/2}) - I}{t} d\mu(s, A) \\ &= \int_{[0,1] \times \mathbb{P}} l_s(X_0^{-1/2} A X_0^{-1/2}) d\mu(s, A), \end{aligned}$$

where we used Theorem 4.2 and that the function $\frac{f_{s,t}(x)-1}{t}$ is in \mathfrak{L} which follows from basic calculations. By this we also have that

$$\int_{[0,1] \times \mathbb{P}} X_0^{1/2} l_s(X_0^{-1/2} A X_0^{-1/2}) X_0^{1/2} d\mu(s, A) = 0.$$

□

Lemma 6.10. *The set $K(\mu)$ is invariant under congruencies i.e., for any $C \in \text{GL}(E)$*

$$CK(\mu)C^* = K(C\mu C^*).$$

Proof. For any $X \in K(\sigma)$ we have

$$\begin{aligned} 0 &= \int_{[0,1] \times \mathbb{P}} l_s(X^{-1/2} A X^{-1/2}) d\mu(s, A) \\ (49) \quad &= \int_{[0,1] \times \mathbb{P}} l_s(X^{-1} A) d\mu(s, A). \end{aligned}$$

Let $C = UP$ the polar decomposition of C i.e., $U^{-1} = U^*$ and $P \in \mathbb{P}$. Then by (49) it follows directly that

$$UK(\mu)U^* = K(U\mu U^*).$$

Similarly we have

$$\begin{aligned} 0 &= P^{-1} \left(\int_{[0,1] \times \mathbb{P}} l_s(X^{-1/2} A X^{-1/2}) d\mu(s, A) \right) P \\ &= \int_{[0,1] \times \mathbb{P}} l_s(P^{-1} X^{-1} A P) d\mu(s, A) \\ &= \int_{[0,1] \times \mathbb{P}} l_s(P^{-1} X^{-1} P^{-1} P A P) d\mu(s, A), \end{aligned}$$

so $PXP \in K(P\mu P)$ i.e., $PK(\mu)P \subseteq K(P\mu P)$. Also then $K(\mu) \subseteq P^{-1}K(P\mu P)P^{-1} \subseteq K(\mu)$ which means

$$PK(\mu)P = K(P\mu P).$$

From this and $UK(\mu)U^* = K(U\mu U^*)$ we get that

$$CK(\mu)C^* = K(C\mu C^*).$$

□

We have already seen that the set $\mathscr{P}([0, 1] \times \mathbb{P})$ is a subset of a Banach space equipped with the total variation norm.

Proposition 6.11. *There exists $\epsilon > 0$ such that for $\mu \in \mathscr{P}([0, 1] \times \mathbb{P})$ with $\text{supp } \mu \subseteq [0, 1] \times \overline{B}_A(\epsilon)$, the equation*

$$\int_{[0,1] \times \mathbb{P}} l_s \left(X^{-1/2} A X^{-1/2} \right) d\mu(s, A) = 0$$

has a unique solution in $\overline{B}_A(\epsilon)$ which is $\Lambda(\mu)$.

Proof. We would like to use the Implicit Function Theorem for Banach spaces. First of all the map

$$F_\mu(X) := F(\mu, X) := \int_{[0,1] \times \mathbb{P}} l_s(X^{-1/2}AX^{-1/2})d\mu(s, A)$$

maps from the product of two Banach spaces $\mathcal{P}([0, 1] \times \mathbb{P}) \times S(E)$ to a Banach space $S(E)$, moreover it is C^∞ . The Fréchet derivative of F_μ is a linear map on $S(E)$. Let $\nu \in \mathcal{P}([0, 1])$ and $\sigma_I \in \mathcal{P}(\mathbb{P})$ such that σ_I is only supported on the singleton $\{I\}$. Then $F_{\nu \times \sigma_I}(X) = \int_{[0,1]} l_s(X^{-1})d\nu(s)$ so $F_{\nu \times \sigma_I}(I) = 0$ and by the property $l'_s(1) = 1$ we have that the Fréchet derivative $DF_{\nu \times \sigma_I}[I] = -id_{S(E)}$. Thus by the Implicit Function Theorem (Theorem 5.9 [21]) there exists an open neighborhood U of $\nu \times \sigma_I$ in $\mathcal{P}([0, 1] \times \mathbb{P})$ and a neighborhood V of $I \in \mathbb{P}$, and a C^∞ mapping $g : U \mapsto V$ such that $F_\mu(X) = 0$ if and only if $X = g(\mu)$ for $\mu \in U$, $X \in V$. Now if we pick $\epsilon > 0$ such that $\overline{B}_I(\epsilon) \subseteq U$ and $\overline{B}_I(\epsilon) \subseteq V$ then the first part of the assertion is proved for $A = I$. The general case for any A follows from Lemma 6.10 with $C = A^{1/2}$.

The second part of the assertion is a consequence of property (8) in Theorem 6.4 which yields that $\Lambda(\mu) \in \overline{B}_A(\epsilon)$ which is then identical to $g(\mu)$ by the first part of the assertion. \square

Corollary 6.12. *Using Lemma 5.5 and Proposition 6.11 it follows that the lambda operator mean $\Lambda(\mu)$ is C^∞ on small enough neighborhoods of $\text{supp } \mu$ assuming that $\text{supp } \mu \subseteq [0, 1] \times \overline{B}_A(\epsilon)$ for small enough $\epsilon > 0$ and fixed $A \in \mathbb{P}$.*

Theorem 6.13. $K(\mu) = \{\Lambda(\mu)\}$ for all $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$.

Proof. We start with a

Claim. *For fixed $X \in \mathbb{P}$ and $t \in (0, 1]$ define*

$$dM_{s,t}(X, \mu)(s, A) := d\mu(s, X^{1/2}f_{s,t}^{-1}(X^{-1/2}AX^{-1/2})X^{1/2})$$

if there exists $Z \in \mathbb{P}$ such that $f_{s,t}(Z) = X^{-1/2}AX^{-1/2}$, in the other case when there exists no such $Z \in \mathbb{P}$ define $dM_{s,t}(X, \mu)(s, A) := 0$. Then the equation

$$(50) \quad X = \Lambda(M_{s,t}(X, \mu))$$

has a unique solution in \mathbb{P} .

Let us briefly explain the reason behind the introduction of $M_{s,t}(X, \mu)$. The above piecewise definition of $M_{s,t}(X, \mu)$ is needed since the inverse map

$$(51) \quad g^{-1}(A) := X^{1/2}f_{s,t}^{-1}(X^{-1/2}AX^{-1/2})X^{1/2}$$

(which is the inverse of $g(A) = M_{s,t}(X, A)$) may not be well defined on the whole of \mathbb{P} . Now the point of our definition is that for fixed $X \in \mathbb{P}$ we

would like to consider the one-parameter family of unique solutions X_u of the equations

$$X_u = \int_{[0,1] \times \mathbb{P}} M_{s,u}(X_u, M_{s,t}(X, A)) d\mu(s, A)$$

for $u \in (0, 1]$ defining the induced operator means in the sense of (42) with respect to the parameter $t \in (0, 1]$. Then we take the limit $u \rightarrow 0+$ to obtain the corresponding lambda operator means depending on the measure $M_{s,t}(X, \mu)$, which itself depends on X, t and a fixed $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$.

Now let us turn to the proof of the Claim. First of all notice that for $X \in \mathbb{P}$ the function $g^{-1}(A)$ defined by (51) is the inverse of $g(A)$, moreover this inverse is well defined and in fact C^∞ since $f_{s,t}(x)$ is an operator monotone function hence analytic and strictly monotone increasing. Indeed, by the above $M_{s,t}(X, \mu)$ is well defined. Moreover for fixed $X \in \mathbb{P}$ the set of all $A \in \mathbb{P}$ such that $M_{s,t}(X, \mu)$ is supported in $[0, 1] \times \{A\}$ is bounded in \mathbb{P} . Now choose a large enough closed set $S \subseteq \mathbb{P}$ such that for all $X, Y \in S$ the support of $M_{s,t}(X, \mu)$ and $M_{s,t}(Y, \mu)$ is included in $[0, 1] \times S$. This is clearly possible since we have the bounds according to Lemma 3.1

$$((1-t) + tx^{-1})^{-1} \leq f_{s,t}(x) \leq (1-t) + tx$$

and can use a similar argument as in the proof of Proposition 5.7. Now choose a large enough $0 < r < \infty$ such that $S \subseteq \overline{B}_I(r)$. Then for all $X, Y \in S$ and for each $A \in \mathbb{P}$ such that $M_{s,t}(X, \mu)$ or $M_{s,t}(Y, \mu)$ is supported in $[0, 1] \times \{A\}$, we have $S \subseteq \overline{B}_A(r)$. By Proposition 5.3 $g_{s,t}(X) = M_{s,t}(X, A)$ is a strict contraction on $\overline{B}_A(r)$, with contraction coefficient

$$\rho = \rho_1 + \frac{\log \frac{be^{-r} + ae^{2r(1-\rho_1)}}{be^{-r} + a}}{2r},$$

where $\rho_1 = \frac{\log \frac{e^{3r(1-t)} + t}{e^r(1-t) + t}}{2r}$ and $a = \frac{s(1-t)}{t+s(1-t)}$, $b = \frac{t}{t+s(1-t)}$. Since $t \leq b \leq 1$ and $0 \leq a \leq 1-t$, by (*), using a similar argument as we did to obtain (31), we get

$$\rho \leq \rho_1 + \frac{\log \frac{te^{-r} + (1-t)e^{2r(1-\rho_1)}}{te^{-r} + (1-t)}}{2r},$$

which is strictly less than 1. Let $f(X) = \Lambda(M_{s,t}(X, \mu))$. Now by property (6) in Theorem 6.4 we have

$$\begin{aligned} & d_\infty(\Lambda(M_{s,t}(X, \mu)), \Lambda(M_{s,t}(Y, \mu))) \\ & \leq \sup_{\mu \text{ is supported on } \{s\} \times \{A\}} \{d_\infty(M_{s,t}(X, A), M_{s,t}(Y, A))\} \\ & \leq \rho d_\infty(X, Y), \end{aligned}$$

i.e., $f(X)$ is a strict contraction on S . Moreover by a similar argument to the proof of Proposition 5.7 we see that $f(S) \subseteq S$. So by Banach's fixed point theorem (50) has a unique solution in S . Since S is an arbitrary large

enough closed subset of \mathbb{P} , it follows that (50) has a unique solution in \mathbb{P} . The claim is proved.

Now let $X \in K(\mu)$. We have that $f_{s,t}(x) = l_s^{-1}(tl_s(x)) \in \mathfrak{m}(t)$ for all $0 < t \leq 1$ and $s \in [0, 1]$. By Lemma 3.1 we have that

$$((1-t) + tx^{-1})^{-1} \leq f_{s,t}(x) \leq (1-t) + tx$$

which means that there exists a small enough $0 < t \leq 1$ such that

$$f_{s,t}(X^{-1/2}AX^{-1/2}) \in \overline{B}_I(\epsilon)$$

for all $A \in \mathbb{P}$ such that μ is supported in $[0, 1] \times \{A\}$. Also it is easy to see due to (38) that

$$\begin{aligned} 0 &= t \left[\int_{[0,1] \times \mathbb{P}} l_s(X^{-1/2}AX^{-1/2}) d\mu(s, A) \right] \\ &= \int_{[0,1] \times \mathbb{P}} l_s(f_{s,t}(X^{-1/2}AX^{-1/2})) d\mu(s, A) \\ &= \int_{[0,1] \times \mathbb{P}} l_s(X^{-1/2}AX^{-1/2}) dM_{s,t}(X, \mu)(s, A) \\ &= \int_{[0,1] \times \mathbb{P}} l_s(A) d \left(X^{-1/2}M_{s,t}(X, \mu)X^{-1/2} \right) (s, A), \end{aligned}$$

in other words

$$(52) \quad 0 = \int_{[0,1] \times \mathbb{P}} l_s(Y^{-1/2}AY^{-1/2}) d \left(X^{-1/2}M_{s,t}(X, \mu)X^{-1/2} \right) (s, A)$$

with $Y = I$. Now by Proposition 6.11 we have that $Y = I$ is the unique solution of (50) on $\overline{B}_I(\epsilon)$ which is $\Lambda(X^{-1/2}M_{s,t}(X, \mu)X^{-1/2})$. Hence from the definition of the lambda operator mean $\Lambda(\mu)$ we have

$$I = \Lambda(X^{-1/2}M_{s,t}(X, \mu)X^{-1/2}).$$

By property (3) in Theorem 6.4 we have that the above is equivalent to

$$\begin{aligned} X &= X^{1/2} \Lambda(X^{-1/2}M_{s,t}(X, \mu)X^{-1/2}) X^{1/2} \\ &= \Lambda(M_{s,t}(X, \mu)). \end{aligned}$$

Now the Claim implies that the solution of the above equation is unique, hence all solutions $X \in K(\mu)$ must be identical. By Theorem 6.9 we have $\Lambda(\mu) \in K(\mu)$, hence $K(\mu) = \{\Lambda(\mu)\}$. \square

Corollary 6.14. *Suppose $\nu \in \mathcal{P}([0, 1])$ and $\sigma \in \mathcal{P}(\mathbb{P})$ and assume that $\sigma(X) := (1-w)\delta_A(X) + w\delta_B(X)$ with $w \in (0, 1)$ where $\delta_A(X)$ denotes the Dirac delta supported on A . Then $\Lambda(\nu \times \sigma)$ is an operator mean in the two variables (A, B) i.e., $\Lambda(\nu \times \sigma) \in \mathfrak{M}$.*

Proof. By Corollary 5.9 and Theorem 6.3 the lambda extension is the strong limit of (induced) operator means i.e.,

$$\Lambda(\nu \times \sigma) = \lim_{t \rightarrow 0+} L_t(\nu \times \sigma).$$

By Lemma 6.1 in [20] the point-wise weak limit of operator means is an operator mean as well, so therefore it follows that the strong limit $\Lambda(\nu \times \sigma)$ of operator means is also an operator mean in the sense of Definition 3.1. \square

Remark 6.2. Theorem 6.13 gives us a tool to solve operator equations that can be written in the form of a generalized Karcher equation

$$\int_{[0,1] \times \mathbb{P}} l_s(A) d\mu(s, A) = 0.$$

The solution can be calculated by choosing a sequence $t_l \rightarrow 0+$ as $l \rightarrow \infty$ and then taking the limit

$$\lim_{l \rightarrow \infty} L_{t_l}(\mu) = \Lambda(\mu).$$

Example 6.1. Consider the function $\log \in \mathfrak{L}$. In [9] formula (V.47) says that

$$\log x = \int_{-\infty}^0 \frac{1}{\lambda - x} - \frac{\lambda}{\lambda^2 + 1} d\lambda$$

which is actually

$$\log x = \int_0^{\infty} \frac{\lambda}{\lambda^2 + 1} - \frac{1}{\lambda + x} d\lambda.$$

By Corollary 2.1 and Corollary 2.2 we get that

$$\log x = \int_{[0,1]} \frac{x - 1}{(1 - s)x + s} d\nu(s)$$

with $d\nu(s) = ds$.

Let $\sigma \in \mathcal{P}(\mathbb{P})$ and $\nu \in \mathcal{P}([0, 1])$. The Dirac delta supported on $\{a\}$ is denoted by $\delta_{\{a\}}(x)$. Consider the generalized Karcher equations

$$\int_{\mathbb{P}} f(X^{-1/2} A X^{-1/2}) d\sigma(A) = 0$$

and their solutions:

$$\begin{aligned} d\nu(s) &= \delta_{\{0\}}(s), \quad f(x) = 1 - x^{-1}, \quad \Lambda(\nu \times \sigma) = \left(\int_{\mathbb{P}} A^{-1} d\sigma(A) \right)^{-1} \\ d\nu(s) &= \delta_{\{1\}}(s), \quad f(x) = x - 1, \quad \Lambda(\nu \times \sigma) = \int_{\mathbb{P}} A d\sigma(A) \\ d\nu(s) &= ds, \quad f(x) = \log x, \quad \Lambda(\nu \times \sigma) = \tilde{\Lambda}(\sigma), \end{aligned}$$

where $\tilde{\Lambda}(\sigma)$ is the original Karcher mean which has been studied first in [24] for positive operators and in the case of positive matrices at many other places, for example see [11, 12, 10, 26, 28, 29]. In particular the first mean is the harmonic mean while the second is the arithmetic mean. Notice that all these cases are extensions of these well known means to the case of measures and operators.

There are other recently found means which are all lambda operator means. The log-determinant α -divergence means found in [14] for positive matrices are also such with ν supported over a singleton $\{s\}$ with $s \in [0, 1]$.

The arithmetic and harmonic means have special properties.

Proposition 6.15. *Let $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$. Then*

$$(53) \quad \left(\int_{\mathbb{P}} A^{-1} d\mu(s, A) \right)^{-1} \leq \Lambda(\mu) \leq \int_{\mathbb{P}} A d\mu(s, A).$$

Proof. First of all by Lemma 3.1 we have for all $s, t \in [0, 1]$ that

$$((1-t) + tx^{-1})^{-1} \leq f_{s,t}(x) \leq (1-t) + tx,$$

in other words

$$[(1-t)X^{-1} + tA^{-1}]^{-1} \leq M_{s,t}(X, A) \leq (1-t)X + tA.$$

Now integrate the above inequality with respect to μ to conclude the assertion using property (4) in Theorem 6.4 with Theorem 6.9 and the previous Example 6.1. \square

One might wonder whether the induced means $L_t(\mu)$ are different from lambda operator means. It turns out that the induced means fit into the overall picture.

Theorem 6.16. *Let $t \in (0, 1]$ and $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$. Then $L_t(\mu)$ is a lambda operator mean, it is the unique solution of the generalized Karcher equation*

$$\int_{[0,1] \times \mathbb{P}} \log_{t+s(1-t)}(X^{-1/2}AX^{-1/2}) d\mu(s, A) = 0.$$

Proof. By definition (41), $L_t(\mu)$ is the unique solution of

$$X = \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu(s, A).$$

So we have

$$\begin{aligned} X &= \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) d\mu(s, A) \\ 0 &= \int_{[0,1] \times \mathbb{P}} M_{s,t}(X, A) - X d\mu(s, A) \\ 0 &= \int_{[0,1] \times \mathbb{P}} \frac{M_{s,t}(X, A) - X}{t} d\mu(s, A) \\ 0 &= \int_{[0,1] \times \mathbb{P}} \frac{f_{s,t}(X^{-1/2}AX^{-1/2}) - I}{t} d\mu(s, A). \end{aligned}$$

Now it is a routine calculation to see that $\frac{f_{s,t}(x)-1}{t} = \log_{t+s(1-t)}(x)$ and $\log_{t+s(1-t)} \in \mathfrak{L}$, since $t + s(1-t) \in [0, 1]$. \square

The above result is a general phenomenon:

Proposition 6.17. *Let $f \in \mathfrak{m}(t)$ for $t \in (0, 1)$. Then $\frac{f(x)-1}{t}$ is in \mathfrak{L} .*

Proof. Simple computation. \square

Corollary 6.18. *Let $\sigma \in \mathcal{P}(\mathbb{P})$ and let $f_t \in \mathfrak{m}(t)$ be the representing function of the operator mean $M_t(X, A)$ with $t \in (0, 1)$. Then*

$$(54) \quad X = \int_{\mathbb{P}} M_t(X, A) d\sigma(A)$$

has a unique solution in \mathbb{P} , moreover it is the unique solution of the generalized Karcher equation of the form

$$\int_{\mathbb{P}} X^{1/2} g \left(X^{-1/2} A X^{-1/2} \right) X^{1/2} d\sigma(A) = 0$$

where $g(x) = \frac{f_t(x)-1}{t}$ is in \mathfrak{L} .

Proof. Simple computation as in the proof of Theorem 6.16. \square

Corollary 6.18 tells us that the matrix power means $P_t(\omega; \mathbb{A})$ defined as the unique positive definite solution of (33) are actually lambda operator means for $t \in [-1, 1]$ i.e., they are unique solutions of generalized Karcher equations. Let us provide the details in the following example.

Example 6.2. For $t \in (0, 1]$, the matrix power means $P_t(\omega; \mathbb{A})$ are the unique positive definite solutions of

$$(55) \quad X = \sum_{i=1}^k w_i X \#_t A_i$$

where

$$A \#_t B = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^t A^{1/2}$$

is the weighted geometric mean of $A, B \in \mathbb{P}$, its representing function is $f_t(x) = x^t$ and $f_t \in \mathfrak{m}(t)$. For $\sigma \in \mathcal{P}(\mathbb{P})$ let us consider a generalized form of (55) in the form

$$(56) \quad X = \int_{\mathbb{P}} X \#_t A d\sigma(A).$$

By Corollary 6.18 this is equivalent to

$$(57) \quad 0 = \int_{\mathbb{P}} X^{1/2} f \left(X^{-1/2} A X^{-1/2} \right) X^{1/2} d\sigma(A)$$

where $f(x) := \frac{x^t-1}{t}$ and $f \in \mathfrak{L}$. Now using the representation V.48 in [9] for the power function x^t , we have that

$$\begin{aligned} f(x) = \frac{x^t-1}{t} &= \frac{\cos \frac{t\pi}{2} - 1}{t} + \int_{-\infty}^0 \left(\frac{1}{\lambda-x} - \frac{\lambda}{\lambda^2+1} \right) |\lambda|^t \frac{\sin(t\pi)}{t\pi} d\lambda \\ &= \frac{\cos \frac{t\pi}{2} - 1}{t} + \int_0^{\infty} \left(\frac{\lambda}{\lambda^2+1} - \frac{1}{\lambda+x} \right) \lambda^t \frac{\sin(t\pi)}{t\pi} d\lambda. \end{aligned}$$

After applying change of variables and using Corollary 2.1 and Corollary 2.2 we get that

$$f(x) = \frac{x^t-1}{t} = \int_0^1 \frac{x-1}{(1-s)x+s} d\nu(s)$$

where

$$d\nu(s) = \frac{s^t}{(1-s)^t} \frac{\sin(t\pi)}{t\pi} ds.$$

Hence by (57) the matrix power means P_t are unique solutions of the generalized Karcher equations

$$(58) \quad 0 = \int_{\mathbb{P}} \int_0^1 X^{1/2} l_s \left(X^{-1/2} A X^{-1/2} \right) X^{1/2} d\nu(s) d\sigma(A).$$

In the finite dimensional case one can prove more by applying Theorem 7.1 in the next section to conclude that the matrix power means are global minimizers of geodesically convex functions. Similar calculations can be carried out for negative values of $t \in [-1, 0)$. Also by taking the limit $t \rightarrow 0$ we obtain the case of the original Karcher mean considered in Example 6.1.

7. CONCLUSION

The previous sections provide the completion of Theorem 2.6 in the case when \mathbb{P} is the finite dimensional cone of positive matrices. Theorem 6.13 tells us that the gradient equations

$$(59) \quad \int_{\mathfrak{L} \times \mathbb{P}} X^{1/2} f(X^{-1/2} A X^{-1/2}) X^{1/2} d\sigma(f, A) = 0$$

given in (22) for $\sigma \in \mathcal{P}(\mathfrak{L} \times \mathbb{P})$ admit unique solutions in \mathbb{P} . Theorem 2.6 shows us that

$$(60) \quad \int_{[0,1] \times \mathbb{P}} X^{1/2} l_s(X^{-1/2} A X^{-1/2}) X^{1/2} d\mu(s, A) = 0$$

has a unique solution and now it is not difficult to see that it is the gradient equation of the critical points of the strictly geodesically convex functional

$$\int_{[0,1] \times \mathbb{P}} LD^s(X, A) d\mu(s, A).$$

This argument essentially can be carried out using Theorem 2.4 and Theorem 2.3. Hence we have the result:

Theorem 7.1. *Let \mathbb{P} be finite dimensional and $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$. Then*

$$(61) \quad \Lambda(\mu) = \arg \min_{X \in \mathbb{P}} \int_{[0,1] \times \mathbb{P}} LD^s(X, A) d\mu(s, A).$$

Remark 7.1. More questions can be asked about the size of the set generated by $\Lambda(\mu)$ for various different $\mu \in \mathcal{P}([0, 1] \times \mathbb{P})$. Perhaps Corollary 6.14 provides the correct setting. How large is the set of two-variable lambda operator means in \mathfrak{M} ? This question is nontrivial and it is very likely that there are operator means in \mathfrak{M} which are not lambda operator means. In particular if $w = 1/2$ in Corollary 6.14 we have some examples when a two variable operator mean is the unique solution of two different generalized Karcher equations. The geometric mean $A \#_{1/2} B = A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2}$ is

such, see [14]. However this redundancy only occurs if $w = 1/2$. This result, though might be important, is out of scope of this paper and will be presented elsewhere. So all in all the set of lambda operator means is of the "size" of \mathfrak{L} if $w \neq 1/2$. In this case we even assumed that the same $f \in \mathfrak{L}$ is considered in the generalized Karcher equation i.e., $\mu = \nu \times \sigma$ is a product. In the general case we expect the set of lambda operator means to be even larger.

ACKNOWLEDGMENT

The author would like to thank the anonymous referee for valuable comments and suggestions. This work was partly supported by the Research Fellowship of the Canon Foundation, SGU project of Kyoto University, the JSPS international research fellowship grant No. 14F04320, the "Lendület" Program (LP2012-46/2012) of the Hungarian Academy of Sciences and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) No. 2015R1A3A2031159.

REFERENCES

- [1] J. Agler, J. E. McCarthy and N. Young, Operator monotone functions and Löwner functions of several variables, *Annals of Math.*, 176:3 (2012), pp. 1783–1826.
- [2] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis: A hitchhiker's guide* (Third ed.). Berlin: Springer, (2006).
- [3] T. Ando, C-K. Li and R. Mathias, Geometric means, *Linear Alg. Appl.*, 385 (2004), pp. 305–334.
- [4] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, Geometric means in a novel vector space structure on symmetric positive-definite matrices, *SIAM. J. Matrix Anal. Appl.* 29 (2006), pp. 328–347.
- [5] F. Barbaresco, Interactions between symmetric cone and information geometries: Bruhat-Tits and Siegel spaces models for higher resolution autoregressive doppler imagery, *Lecture Notes in Computer Science* 5416 (2009), pp. 124–163.
- [6] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, Riemannian geometry applied to BCI classification, preprint.
- [7] A. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*, Birkhäuser, Basel, (2003).
- [8] J. Bendat and S. Sherman, Monotone and convex operator functions, *Tran. of the AMS* 79 (1955), pp. 58–71.
- [9] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, (1996).
- [10] R. Bhatia, *Positive Definite Matrices*, Princeton University Press, Princeton, New Jersey (2007), ISBN: 978-0-691-12918-1.
- [11] R. Bhatia and J. Holbrook, Riemannian geometry and matrix geometric means, *Linear Algebra Appl.* 413 (2006), pp. 594–618.
- [12] R. Bhatia and R. Karandikar, Monotonicity of the matrix geometric mean, to appear in *Math. Ann.*
- [13] M. R. Bridson and A. Häflicher, *Metric Spaces of Non-Positive Curvature*, 643 p. Springer-Verlag, Berlin Heidelberg (1999), ISBN: 978-3-540-64324-1.
- [14] Z. Chebbi and M. Moakher, Means of Hermitian positive-definite matrices based on the log-determinant α -divergence function, *Linear Alg. Appl.*, 436 (2012), pp. 1872–1889.

- [15] P. Fillard, V. Arsigny, X. Pennec and N. Ayache, Joint estimation and smoothing of clinical DT-MRI with a Log-Euclidean metric, Research Report RR-5607, INRIA, Sophia-Antipolis, France, June 2005.
- [16] F. Hansen, The fast track to Löwner's theorem, *Linear Alg. Appl.*, 438 (2013), pp. 4557–4571.
- [17] R. Kadison, Strong continuity of operator functions, *Pacific J. Math.* 26 (1968), pp. 121–129.
- [18] H. Karcher, Riemannian center of mass and mollifier smoothing, *Comm. Pure Appl. Math.*, vol. 30 (1977), pp. 509–541.
- [19] H. Karcher, Riemannian comparison constructions, *Global Differential Geometry*, MAA Studies in Math., Vol. 27, S.-S. Chern, editor, MAA (1989), pp. 170–222.
- [20] F. Kubo and T. Ando, Means of positive linear operators, *Math. Ann.*, 246 (1980), pp. 205–224.
- [21] S. Lang, *Fundamentals of Differential Geometry*, Graduate Texts in Mathematics, Springer, (1999).
- [22] G. Laroitonda, Nonpositive curvature: a geometrical approach to Hilbert-Schmidt operators, *Differential Geom. Appl.* 25:6 (2007), pp. 679–700.
- [23] J. Lawson and Y. Lim, A general framework for extending means to higher orders, *Colloq. Math.*, 113 (2008), pp. 191–221.
- [24] J. Lawson and Y. Lim, Karcher means and Karcher equations of positive definite operators, *Tran. of the AMS Series B* 1 (2012), pp. 1–22.
- [25] J. Lawson and Y. Lim, Weighted means and Karcher equations of positive operators, *Proceedings of the National Academy of Sciences*, vol. 110 no. 39 (2013), pp. 15626–15632.
- [26] J. Lawson and Y. Lim, Monotonic properties of the least squares mean, *Math. Ann.* (2010), doi:10.1007/s00208-010-0603-6.
- [27] J. Lawson and Y. Lim, A Lipschitz constant formula for vector addition in cones with applications to Stein-like equations, *Positivity* 16 (2012), pp. 81–95.
- [28] Y. Lim and M. Pálfi, The matrix power means and the Karcher mean, *J. Func. Anal.*, Vol. 262, No. 4 (2012), pp. 1498–1514.
- [29] M. Moakher, A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices, *SIAM J. Matrix Anal. Appl.*, 26 (2005), pp. 735–747.
- [30] K. Musiał, Pettis integral, *Handbook of Measure Theory I*, Elsevier, Amsterdam (2002), pp. 531–586.
- [31] K.-H. Neeb, Compressions of infinite-dimensional bounded symmetric domains, *Semigroup Forum* 61 (2001), 71–105.
- [32] M. Pálfi, A Multivariable Extension of Two-variable Matrix Means, *SIAM J. Matrix Anal. Appl.*, 32 (2011), pp. 385–393.
- [33] M. Pálfi, Means in metric spaces and the center of mass, *J. Math. Anal. Appl.*, 381 (2011), pp. 383–391.
- [34] M. Pálfi, Weighted matrix means and symmetrization procedures, *Linear Alg. Appl.* (2011), doi:10.1016/j.laa.2011.06.036.
- [35] A. Papadopoulos, *Metric Spaces, Convexity and Nonpositive Curvature*, IRMA Lect. Math. Theor. Phys., vol. 6, European Mathematical Society (EMS), Zrich, 2005.
- [36] W. Rudin, *Functional Analysis*, Prentice-Hall, McGraw-Hill, 1973.
- [37] M. Spivak, *A comprehensive introduction to differential geometry Vol. 2*, Publish or Perish, Houston, Texas (1979), ISBN: 0-914098-88-8.
- [38] R. L. Schilling, R. Song, Z. Vondraček, *Bernstein Functions: Theory and Applications*, de Gruyter Studies in Mathematics 37, Springer, Berlin, 2010.
- [39] K.-T. Sturm, Probability measures on metric spaces of nonpositive curvature, in: *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, Eds. P. Auscher et. al., *Contemp. Math.* 338, Amer. Math. Soc. (AMS), Providence, 2003.

- [40] A.C. Thompson, On certain contraction mappings in a partially ordered vector space, Proc. Amer. Math. Soc., 14 (1963), pp. 438–443.
- [41] J. Weidman, Linear Operators in Hilbert Spaces, Springer Verlag, 1980.

DEPARTMENT OF MATHEMATICS, KYOTO UNIVERSITY, KYOTO 606-8502, JAPAN.

E-mail address: `palfia.miklos@aut.bme.hu`